



PT muscle cells in vivo - useful for treating patients with muscular disorders, e.g. Duchenne and Becker muscular dystrophy, myopathy and cardiac diseases

PS Example 2; Page 12; 35pp; English.

CC The present sequence represents a muscle-specific peptide that specifically binds to muscle cells in vivo. The present invention also describes: (1) determination of muscle-specific peptides, comprising constructing a bacteriophage library that expresses random peptides at the amino terminus of a phage protein, and hybridizing the resulting phage to particular muscle cells (the peptides are selected based upon in vivo binding); (2) targeting a pharmaceutical composition to muscle cells by ligating a muscle-specific peptide to the composition; and (3) a nucleotide sequence encoding all muscle-specific peptides. The muscle-specific peptides form compositions including pharmaceutical compositions that are useful in gene therapy methods in treating patients with muscular disorders. These disorders include Duchenne muscular dystrophy (DMD), Becker muscular dystrophy (BMD), myopathies (including myotic dystrophy, myotonia congenita, dermatomyositis, polymyositis and polynuropathy) and cardiac diseases. The muscle-specific peptides allow the specific targeting of agents used to treat muscular conditions by overcoming the current problems associated with gene therapy. These include the impractical nature of widespread intramuscular transduction in a clinical context, the inaccessibility of certain muscles (e.g. heart and diaphragm), the mode of administration, and the capacity of the vector used.

CC Sequence 7 AA;

CC Query Match 100.0%; Score 36; DB 20; Length 7;

CC Best Local Similarity 100.0%; Pred. No. 7.8e+05; Indels 0; Gaps 0;

CC Matches 7; Conservative 0; Mismatches 0;

CC 1 GETRAPL 7

CC 1 GETRAPL 7

CC AAB03032 standard; peptide; 7 AA.

CC AAB03032;

CC 25-SEP-2000 (first entry)

CC MN protein CA domain-binding peptide, SEQ ID NO:107.

CC MN protein; tumour associated cell adhesion molecule; oncoprotein; proteoglycan domain; PG domain; carbonic anhydrase; CA domain; abnormal expression; neoplastic disease; cancer; gene therapy; phage display library.

CC Synthetic.

CC WO200024913-A2.

CC 04-MAY-2000.

CC 22-OCT-1999; 99WO-US24879.

CC 23-OCT-1998; 98US-0177776.

CC 23-OCT-1998; 98US-0178115.

CC (FARB ) BAYER CORP.

CC (VIRO-) INST VIROLOGY.

CC Zavada J, Pastorekova S, Pastorek J;

CC MPI; 2000-350752/30.

CC A molecule which specifically binds to a site on MN protein

PT (oncoprotein) and prevents adhesion of vertebrate cells to the protein, useful for treating preneoplastic or neoplastic diseases such as cancer

PS Claim 5; Page 71; 154pp; English.

CC The invention relates to the inhibition of cell adhesion mediated by the MN oncoprotein (also known as the MN/CA IX isoenzyme or the MN/5250 protein). The MN protein is a tumour-associated adhesion molecule which comprises a proteoglycan-like (PG) domain (AAB03017) which contains the protein's binding site, and a carbonic anhydrase (CA) domain (AAB03018). Abnormal expression of the MN protein is associated with tumorigenicity. The invention encompasses molecules (e.g., proteins and peptides) which which specifically bind to a site on the MN protein, thereby preventing adhesion of vertebrate cells to the protein in a cell adhesion assay. It also encompasses MN proteins or MN protein fragments which can be added to the extracellular environment to prevent the adhesion of vertebrate cells to each other. The invention also relates to the identification of the binding site of the MN protein and to a method of identifying a site on an MN protein to which cells adhere, comprising testing a series of overlapping peptides from the protein in a cell adhesion assay. The invention encompasses a vector comprising an expression control sequence operatively linked to a nucleic acid encoding the variable domain of a MN-specific antibody, where the domains are separated by a flexible linker peptide (AAB03035) and the vector inhibits the growth of a vertebrate preneoplastic or neoplastic cell that abnormally expresses MN protein. The invention also encompasses a vector comprising a nucleic acid encoding a cytotoxic protein or peptide operatively linked to the MN gene promoter, which inhibits the growth of a vertebrate preneoplastic or neoplastic cell. Also claimed is a repressor complex that binds to the MN gene promoter (AAB52473). MN proteins and peptides, MN-binding proteins and peptides, and expression vectors encoding such proteins and peptides are useful for treating patients with preneoplastic or neoplastic disease (e.g., cancers) associated with or characterised by abnormal MN expression. Sequences AAB03032-B03034 and AAB03055-B03058 represent synthetic phage display library peptides which bind to the CA domain of the human MN protein (AAB03005).

CC Sequence 7 AA;

CC Query Match 100.0%; Score 36; DB 21; Length 7;

CC Best Local Similarity 100.0%; Pred. No. 7.8e+05; Indels 0; Gaps 0;

CC Matches 7; Conservative 0; Mismatches 0;

CC 1 GETRAPL 7

CC 1 GETRAPL 7

CC AAY84992 standard; peptide; 7 AA.

CC AAY84992;

CC 21-AUG-2000 (first entry)

CC Amino acid sequence of an endothelial cell-binding peptide motif.

CC Endothelial cell-binding peptide; Genetic display package; peptide display library; affinity selection; population display package; cell proliferation; cell differentiation; cell death; cell migration; angiogenic activity; infective peptide; anti-fungal; anti-bacterial; receptor protein effector.

CC Unidentified.

CC WO200023465-A2.

CC 27-APR-2000.

CC 19-OCT-1999; 99WO-US24276.

PR 19-OCT-1998; 98US-0174943.  
 XX  
 XX (MITO-) MITOTIX INC.  
 XX  
 XX Gyuris J, Morris AJ;  
 PI  
 XX WPI; 2000-339649/29.  
 DR  
 XX  
 XX  
 PT Generating a peptide with a selected biological activity useful for  
 PT identifying endothelial inhibitors and peptides with anti-angiogenic  
 PT activity by combining peptide display libraries in a display and a  
 PT secretion mode -  
 XX  
 XX  
 PS Disclosure; Page 42; 86pp; English.  
 XX  
 XX  
 CC The present sequence represents an endothelial cell-binding peptide  
 CC motif. The peptide may be identified using the method of the invention.  
 CC The specification describes a method for generating a peptide having a  
 CC selected biological activity. The method comprises displaying the  
 CC peptides on the outer surface of a genetic display package to create a  
 CC peptide display library, and using affinity selection to enrich the  
 CC population display packages for those containing peptides which have  
 CC desired specificity to the target cell. The method may be used in the  
 CC selection of peptides having effects on cell proliferation,  
 CC differentiation, death and migration, as well as in the identification  
 CC of peptides which have anti-proliferative activity with respect to one  
 CC or more types of cells, peptides with (anti-)angiogenic activity,  
 CC anti-infective peptides (e.g. which are active as anti-fungal or  
 CC anti-bacterial), receptor protein effectors, and ligands for orphan  
 CC receptors for which no ligand is known. Moreover, the method may be  
 CC used to test functional ligand-receptor or ligand-ion channel  
 CC interactions for cell surface-localized receptors and channels.  
 XX  
 SO Sequence 7 AA;  
 Query Match 100.0%; Score 36; DB 21; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 7.8e+05;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GETRAPL 7  
 DB 1 GETRAPL 7  
 RESULT 4  
 AAY92734  
 ID AAY92734 standard; peptide; 7 AA.  
 AC AAY92734;  
 XX  
 XX 29-AUG-2000 (first entry)  
 DT  
 XX  
 XX Heptapeptide 4 mimotope of surface LOS of serogroup B meningococcus.  
 DE Heptapeptide; epitope; mimotope; surface lipo-oligosaccharide; LOS;  
 XX  
 XX meningitis; anti-bacterial; anti-inflammatory; vaccine.  
 KW  
 XX  
 OS Neisseria meningitidis.  
 XX  
 XX WO200025814-A2.  
 PN  
 XX 11-MAY-2000.  
 PD  
 XX 27-OCT-1999; 99WO-GB03559.  
 PF  
 XX 30-OCT-1998; 98GB-0023835.  
 PR  
 XX  
 XX (UNLO ) UNIV COLLEGE LONDON.  
 PA (NABI-) NAT INST BIOLOGICAL STANDARDS & CO.  
 XX  
 XX Charalambous BM, Feavers IM;  
 PI  
 XX WPI; 2000-365403/31.

XX  
 XX Use of a mimotope of a surface lipooligosaccharide of a serogroup B  
 PT meningococcus for a vaccine against serogroup B meningococci  
 PT  
 XX  
 XX Claim 6; Page 25; 39pp; English.  
 PS  
 XX  
 XX AAY92731-66 are heptapeptide epitopes which are mimotopes of a surface  
 CC lipooligosaccharide (LOS) of a serogroup B meningococcus. The mimotopes  
 CC are useful in vaccines against serogroup B meningococci, especially  
 CC Neisseria meningitidis, which causes meningitis. The LOS is a surface  
 CC glycolipid that forms a major outer membrane component and possesses a  
 CC terminal galactose acceptor sites for sialic acid. Sialylation of the  
 CC LOS in immunotype B meningococci may enhance the ability of the organism  
 CC to evade the human immune response. Therefore alternative target antigens  
 CC on the surface of serogroup B meningococci are important.  
 XX  
 SO Sequence 7 AA;  
 Query Match 100.0%; Score 36; DB 21; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 7.8e+05;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GETRAPL 7  
 DB 1 GETRAPL 7  
 RESULT 5  
 AAY53711  
 ID AAY53711 standard; peptide; 7 AA.  
 AC AAY53711;  
 XX  
 XX 22-FEB-2000 (first entry)  
 DT  
 XX  
 XX Amino acid sequence of a synthetic mimotope of HCV.  
 DE  
 XX  
 XX Mimotope; antigen; HCV; conformational epitope; HCV antibody; vaccine;  
 KW chronic HCV infection.  
 XX  
 XX Synthetic.  
 OS Hepatitis C virus.  
 XX  
 XX WO958561-A1.  
 PN  
 XX 18-NOV-1999.  
 PD  
 XX 14-MAY-1999; 99WO-FR01155.  
 PF  
 XX 14-MAY-1998; 98FR-0006335.  
 PR  
 XX (INMR ) PASTEUR MERIEUX SERUMS & VACCINS SA.  
 PA  
 XX Barban V;  
 PI  
 XX WPI; 2000-053080/04.  
 DR  
 XX  
 XX New peptides useful for prophylactic and therapeutic treatment of  
 PT hepatitis C -  
 PT  
 XX  
 XX Claim 1; Page 22; 45pp; French.  
 PS  
 XX  
 XX AAY53708-14 represent a synthetic mimotopes of an antigen of Hepatitis C  
 CC virus (HCV). The peptides comprise an amino acid sequence that imitates  
 CC a conformational epitope of the antigenic structure of the virus without  
 CC wholly corresponding to an amino acid sequence of this antigen. The  
 CC peptides are capable of reacting with a specific antibody against a HCV  
 CC antigen. The peptides are used for the prophylactic and therapeutic  
 CC treatment and prevention of HCV, especially for use as a vaccine. The  
 CC peptides are also useful as reagents for the diagnosis of HCV and/or the  
 CC susceptibility to chronic infection by HCV, comprising the determination  
 CC of a humoral and/or specific cellular mediated response of the peptides  
 CC on a whole blood sample.

XX Sequence 7 AA;

Query Match 100.0%; Score 36; DB 21; Length 7;  
Best Local Similarity 100.0%; Pred. No. 7.8e+05;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GETRAPL 7  
Db 1 GETRAPL 7

RESULT 6  
ABP47722

ID ABP47722 standard; Peptide: 7 AA.

AC ABP47722;

XX 19-AUG-2002 (first entry)

DE N. meningitidis LOS peptidic mimotope related peptide #87.

KW Neisseria meningitidis; meningococcus; meningococcal; vaccine; LOS;  
KW lipooligosaccharide; monoclonal antibody; antibacterial; infection;  
KM antiinflammatory; meningococcal disease.

XX Neisseria meningitidis.

OS Synthetic.

XX WO200228688-A2.

XX 11-APR-2002.

PF 03-OCT-2001; 2001MO-EP11409.

PR 03-OCT-2000; 2000GB-0024200.

PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.

PI De Bolle XT, Letesson J, Lobet Y, Mertens PY, Poolman J, Voet P;

DR WPI; 2002-479596/51.

XX Novel mimotope of Neisseria meningitidis surface, for treating  
PT meningococcal disease, comprising a peptide epitope obtainable by  
screening peptide library with a specific monoclonal antibody -

XX Claim 25; Page 51; 55pp; English.

XX The present invention describes mimotopes (I) of a surface U3, 7, 9,  
CC of meningococcal lipopolysaccharides (LOS) of Neisseria meningitidis  
CC comprising a peptide epitope obtained by screening a peptide library  
CC with a monoclonal antibody (Mab) like 4B12C10, H44/24, H44/58, H44/70  
CC or H44/78. (I) is antigenically cross-reactive with Mab. (I) have  
CC antibacterial and antiinflammatory activities, and can be used in  
CC vaccines. Mab is useful in the identification of (I). (I) or Mab are  
CC useful as a medicament, and also in the manufacture of a medicament for  
CC treating or preventing meningococcal disease. (I) and Mab are useful  
CC for treating a patient suffering from or susceptible to meningococcal  
CC disease by administering (I) or Mab to the patient. (I) is useful in  
CC a diagnostic assay for meningococcal infection to detect antibodies  
CC against U3, 7, 9, LOS and to detect the presence of U3, 7, 9 immunotype  
CC meningococcus in a sample from a patient. ABN8464 to ABN8467 and  
CC ABP4736 to ABP4754 represent sequences used in the exemplification  
CC of the present invention.

XX Sequence 7 AA;

Query Match 100.0%; Score 36; DB 23; Length 7;  
Best Local Similarity 100.0%; Pred. No. 7.8e+05;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GETRAPL 7

Db 1 GETRAPL 7

RESULT 7  
ABG00023  
ID ABG00023 standard; Protein; 357 AA.

XX ABG00023;

XX 13-FEB-2002 (first entry)

DE Novel human diagnostic protein #14.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.

XX Homo sapiens.

XX WO200175067-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001MO-US08631.

XX 31-MAR-2000; 2000US-0540217.

XX 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

XX N-PSDB; AAS64210.

XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -

XX Claim 20; SEQ ID No 30382; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG0377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 357 AA;

Query Match 94.4%; Score 34; DB 22; Length 357;  
Best Local Similarity 85.7%; Pred. No. 31;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GETRAPL 7  
Db 122 GETRAPM 128



```

RESULT 8
AAU64543
ID AAU64543 standard; Protein; 179 AA.
XX
XX AAU64543;
AC
XX
XX 27-FEB-2002 (first entry)
DT
XX
DE Propionibacterium acnes immunogenic protein #25439.
XX
XX SAPHO syndrome; synovitis; acne; pustulosis; hyperostosis; osteomyelitis;
XX uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
XX inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
XX dermatological; osteopathic; neuroprotectant.
XX
XX Propionibacterium acnes.
OS
XX
XX WO200181581-A2.
PN
XX
XX 01-NOV-2001.
PD
XX
XX 20-APR-2001; 2001WO-US12865.
PF
XX
XX 21-APR-2000; 2000US-199047P.
PR 02-JUN-2000; 2000US-208841P.
PR 07-JUL-2000; 2000US-216747P.
XX
XX (CORI-) CORIXA CORP.
PA
XX
XX Skeiky YAM, Persing DH, Mitcham JL, Wang SS, Bhactia A;
XX L'maisonneuve J, Zhang Y, Jen S, Carter D;
XX
XX MPI; 2001-616774/71.
DR N-PSDB; AAS59645.
XX
XX Propionibacterium acnes polypeptides and nucleic acids useful for
XX vaccinating against and diagnosing infections, especially useful for
XX treating acne vulgaris -
XX
XX Example 1; SEQ ID NO 25738; 1069pp; English.
XX
XX Sequences AAU9105-AAU68017 represent Propionibacterium acnes immunogenic
XX polypeptides. The proteins and their associated DNA sequences are used in
XX the treatment, prevention and diagnosis of medical conditions caused by
XX P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
XX pustulosis, hyperostosis and osteomyelitis), uveitis and endophthalmitis.
XX P. acnes is also involved in infections of bone, joints and the central
XX nervous system, however it is particularly involved in the inflammatory
XX lesions associated with acne vulgaris. A method for detecting the
XX presence or absence of P. acnes in a patient comprises contacting a
XX sample with a binding agent that binds to the proteins of the invention
XX and determining the amount of bound protein in the sample. The
XX polypeptides may be used as antigens in the production of antibodies
XX specific for P. acnes proteins. These antibodies can be used to
XX downregulate expression and activity of P. acnes polypeptides and
XX therefore treat P. acnes infections. The antibodies may also be used as
XX diagnostic agents for determining P. acnes presence, for example, by
XX enzyme linked immunosorbent assay (ELISA).
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 179 AA;
SQ

```

```

Query Match          91.7%; Score 33; DB 22; Length 179;
Best Local Similarity 85.7%; Pred. No. 24;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 GETRABL 7
    |||:|
DB 84 GETRSP 90

```

```

RESULT 9
ABB59631
ID ABB59631 standard; Protein; 949 AA.
XX
XX ABB59631;
AC
XX
XX 26-MAR-2002 (first entry)
DT
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 5685.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical.
XX
XX Drosophila melanogaster.
OS
XX
XX WO200171042-A2.
EN
XX
XX 27-SEP-2001.
PD
XX
XX 23-MAR-2001; 2001WO-US09231.
PF
XX
XX 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE ) PE CORP NY.
PA
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX
XX MPI; 2001-656860/75.
DR N-PSDB; ABL03734.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -
XX
XX Disclosure; SEQ ID NO 5685; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABU16176-ABU30511), expressed DNA
XX sequences (ABU1840-ABU16175) and the encoded proteins
XX (AAB57737-ABB72072).
XX
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 949 AA;
SQ

```

```

Query Match          91.7%; Score 33; DB 22; Length 949;
Best Local Similarity 85.7%; Pred. No. 15+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 GETRABL 7
    |||:|
DB 236 GETRSP 242

```

```

RESULT 10
ABB59856
ID ABB59856 standard; Protein; 949 AA.
XX
XX ABB59856;
AC
XX
XX 26-MAR-2002 (first entry)
DT
XX
XX Drosophila melanogaster polypeptide SEQ ID NO 6360.
DE Drosophila; developmental biology; cell signalling; insecticide;
XX
XX

```

```

QY 1 GETRABL 7
    |||:|
DB 84 GETRSP 90

```

KM pharmaceutical.  
XX  
OS Drosophila melanogaster.  
XX  
PN WO200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001WO-US09221.  
XX  
XX 23-MAR-2000; 2000US-191637P.  
PR 11-JUL-2000; 2000US-0614150.  
XX  
PA (PEKE ) PE CORP NY.  
XX  
PI Venter JC, Adams M, Li PWD, Myers EW,  
XX  
XX WPI; 2001-656860/75.  
DR N-PSDB; ABL03959.  
XX  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
PS Disclosure; SEQ ID NO 6360; 21pp + Sequence Listing; English.  
XX  
XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA  
CC sequences (AB101840-AB16175) and the encoded proteins  
CC (ABBS7737-ABBS72072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at fcp.wipo.int/pub/published\_pct\_sequences.  
CC  
XX  
SQ Sequence 949 AA;  
  
Query Match 91.7%; Score 33; DB 22; Length 949;  
Best Local Similarity 85.7%; Pred. No. 1.5e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 GETRAPL 7  
DB 236 GETRSPL 242  
  
RESULT 11  
AAG61717  
ID AAG61717 standard; Protein; 113 AA.  
XX  
XX AAG61717;  
XX  
XX 18-OCT-2000 (first entry)  
XX  
XX Arabidopsis thaliana protein fragment SEQ ID NO: 80115.  
DE  
XX  
XX Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KM termination sequence.  
XX  
XX Arabidopsis thaliana.  
OS  
XX  
XX EP1033405-A2.  
PN  
XX  
XX 06-SEP-2000.  
PD  
XX 25-FEB-2000; 2000EP-0301439.  
PF  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR

PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130691.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 11-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 14-MAY-1999; 99US-0134768.  
PR 18-MAY-1999; 99US-0134941.  
PR 19-MAY-1999; 99US-0135124.  
PR 20-MAY-1999; 99US-0135353.  
PR 21-MAY-1999; 99US-0135829.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 24-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140635.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.

```
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144844.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145813.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145919.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147392.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149375.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151338.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.

PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158332.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 18-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0162193.
PR 29-OCT-1999; 99US-0162142.

Query Match 88.9%; Score 32; DB 21; Length 113;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GENTRAP 6
DB 91 GENTRAP 96

RESULT 12
AAG04903
ID AAG04903 standard; Protein; 242 AA.
XX
AC AAG04903;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 1106.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
```

PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-01300449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136021.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140355.  
PR 23-JUN-1999; 99US-0140355.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141844.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143544.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.

PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 21-JUL-1999; 99US-0144684.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 25-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 05-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150684.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158359.  
PR 13-OCT-1999; 99US-0159283.  
PR 13-OCT-1999; 99US-0159284.  
PR 13-OCT-1999; 99US-0159295.

PR	30-APR-1999	99US-0132407
PR	04-MAY-1999	99US-0132484
PR	05-MAY-1999	99US-0132485
PR	06-MAY-1999	99US-0132486
PR	06-MAY-1999	99US-0132487
PR	07-MAY-1999	99US-0132488
PR	11-MAY-1999	99US-0132489
PR	14-MAY-1999	99US-0134226
PR	14-MAY-1999	99US-0134219
PR	14-MAY-1999	99US-0134221
PR	18-MAY-1999	99US-0134370
PR	18-MAY-1999	99US-0134378
PR	19-MAY-1999	99US-0134991
PR	20-MAY-1999	99US-0135124
PR	21-MAY-1999	99US-0135353
PR	24-MAY-1999	99US-0135629
PR	25-MAY-1999	99US-0136021
PR	27-MAY-1999	99US-0136382
PR	28-MAY-1999	99US-0136782
PR	01-JUN-1999	99US-0137222
PR	04-JUN-1999	99US-0137528
PR	04-JUN-1999	99US-0137502
PR	07-JUN-1999	99US-0138074
PR	08-JUN-1999	99US-0138094
PR	10-JUN-1999	99US-0138540
PR	10-JUN-1999	99US-0138847
PR	14-JUN-1999	99US-0139119
PR	16-JUN-1999	99US-0139452
PR	16-JUN-1999	99US-0139453
PR	17-JUN-1999	99US-0139482
PR	18-JUN-1999	99US-0139484
PR	18-JUN-1999	99US-0139485
PR	18-JUN-1999	99US-0139486
PR	18-JUN-1999	99US-0139487
PR	18-JUN-1999	99US-0139488
PR	18-JUN-1999	99US-0139489
PR	18-JUN-1999	99US-0139490
PR	18-JUN-1999	99US-0139491
PR	18-JUN-1999	99US-0139492
PR	18-JUN-1999	99US-0139493
PR	18-JUN-1999	99US-0139494
PR	18-JUN-1999	99US-0139495
PR	18-JUN-1999	99US-0139496
PR	18-JUN-1999	99US-0139497
PR	18-JUN-1999	99US-0139498
PR	18-JUN-1999	99US-0139499
PR	18-JUN-1999	99US-0139500
PR	18-JUN-1999	99US-0139501
PR	18-JUN-1999	99US-0139502
PR	18-JUN-1999	99US-0139503
PR	18-JUN-1999	99US-0139504
PR	18-JUN-1999	99US-0139505
PR	18-JUN-1999	99US-0139506
PR	18-JUN-1999	99US-0139507
PR	18-JUN-1999	99US-0139508
PR	18-JUN-1999	99US-0139509
PR	18-JUN-1999	99US-0139510
PR	18-JUN-1999	99US-0139511
PR	18-JUN-1999	99US-0139512
PR	18-JUN-1999	99US-0139513
PR	18-JUN-1999	99US-0139514
PR	18-JUN-1999	99US-0139515
PR	18-JUN-1999	99US-0139516
PR	18-JUN-1999	99US-0139517
PR	18-JUN-1999	99US-0139518
PR	18-JUN-1999	99US-0139519
PR	18-JUN-1999	99US-0139520
PR	18-JUN-1999	99US-0139521
PR	18-JUN-1999	99US-0139522
PR	18-JUN-1999	99US-0139523
PR	18-JUN-1999	99US-0139524
PR	18-JUN-1999	99US-0139525
PR	18-JUN-1999	99US-0139526
PR	18-JUN-1999	99US-0139527
PR	18-JUN-1999	99US-0139528
PR	18-JUN-1999	99US-0139529
PR	18-JUN-1999	99US-0139530
PR	18-JUN-1999	99US-0139531
PR	18-JUN-1999	99US-0139532
PR	18-JUN-1999	99US-0139533
PR	18-JUN-1999	99US-0139534
PR	18-JUN-1999	99US-0139535
PR	18-JUN-1999	99US-0139536
PR	18-JUN-1999	99US-0139537
PR	18-JUN-1999	99US-0139538
PR	18-JUN-1999	99US-0139539
PR	18-JUN-1999	99US-0139540
PR	18-JUN-1999	99US-0139541
PR	18-JUN-1999	99US-0139542
PR	18-JUN-1999	99US-0139543
PR	18-JUN-1999	99US-0139544
PR	18-JUN-1999	99US-0139545
PR	18-JUN-1999	99US-0139546
PR	18-JUN-1999	99US-0139547
PR	18-JUN-1999	99US-0139548
PR	18-JUN-1999	99US-0139549
PR	18-JUN-1999	99US-0139550
PR	18-JUN-1999	99US-0139551
PR	18-JUN-1999	99US-0139552
PR	18-JUN-1999	99US-0139553
PR	18-JUN-1999	99US-0139554
PR	18-JUN-1999	99US-0139555
PR	18-J	

PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 25-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 28-JUL-1999; 99US-0145919.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 10-AUG-1999; 99US-0147935.  
PR 11-AUG-1999; 99US-0148171.  
PR 12-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157865.  
PR 06-OCT-1999; 99US-0158029.  
PR 07-OCT-1999; 99US-0158232.  
PR 08-OCT-1999; 99US-0158369.  
PR 12-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159337.  
PR 14-OCT-1999; 99US-0159637.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.

PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161922.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 88.9%; Score 32; DB 21; Length 242;  
Best Local Similarity 100.0%; Pred. No. 57;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GETRAP 6  
Db 220 GETRAP 225

RESULT 14  
AAB54093  
ID AAB54093 standard; Protein; 271 AA.  
XX  
AC AAB54093;  
XX  
DT 09-MAR-2001 (first entry)  
XX  
DE Human pancreatic cancer antigen protein sequence SEQ ID NO:545.  
XX  
KW Human; pancreas; pancreatic cancer; pancreatic cancer antigen;  
KW detection; diagnosis; identification; cytostatic; neuroprotective;  
KW neotrophic; immunomodulatory; relaxant; contraceptive; gynaecological;  
KW antiinflammatory; cardiant; gene therapy; chromosome mapping;  
KW linkage analysis; tissue identification; tissue typing; forensic;  
KW neural; immune system; muscular; reproductive; gastrointestinal;  
KW palmonary; cardiovascular; renal; proliferative.  
XX  
XX Homo sapiens.  
OS  
XX  
XX WO200055320-A1.  
XX  
PD 21-SEP-2000.  
XX  
XX 08-MAR-2000; 2000MO-US05989.  
PF  
XX 12-MAR-1999; 99US-0124270.  
PR  
XX (HUMA-) HUMAN GENOME SCI INC.  
PA  
XX Rosen CA, Ruben SM,  
XX WPI; 2000-579444/54.  
XX N-PSDB; AAC98858.  
DR  
XX New nucleic acid that is a pancreatic cancer antigen for preventing,  
PT treating, or ameliorating a medical condition, particular pancreatic  
PT cancer, or for use in assays for diagnosing a pathological condition -  
PS Claim 11; Page 980-981; 1379pp; English.  
XX  
XX AAC98773 to AAC99231 encode the human pancreatic cancer associated  
CC proteins, called pancreatic cancer antigens, given in AAB54008 to  
CC AAB54466. The human pancreatic cancer antigens have cytostatic,  
CC neuroprotective, neotrophic, immunomodulatory, relaxant, contraceptive,

CC gynaecological, cardiant and antiinflammatory activities, and can be used  
 CC in gene therapy. The polynucleotide and proteins can be used for  
 CC preventing, treating, or ameliorating a medical condition or in assays  
 CC for diagnosing a pathological condition or a susceptibility to one in a  
 CC subject. Binding partners to the proteins and the activity of the  
 CC proteins can be identified. The pancreatic cancer antigens can be used to  
 CC detect, treat or prevent pancreatic disorders, especially cancer.  
 CC Agonists and antagonists to the antigens can be screened for. The  
 CC pancreatic cancer antigen polynucleotides can be used to design nucleic  
 CC acid hybridisation probes that can be used in chromosome mapping, linkage  
 CC analysis, tissue identification and/or typing and a variety of forensic  
 CC and diagnostic methods. The proteins can be used to generate antibodies  
 CC which are used to purify, detect and target the polypeptides, including  
 CC both in vivo and in vitro diagnostic and therapeutic methods. The  
 CC proteins can be used to treat or prevent neural, immune system, muscular,  
 CC reproductive, gastrointestinal, pulmonary, cardiovascular, renal or  
 CC proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent  
 CC sequences used in the exemplification of the present invention.

SQ Sequence 271 AA;

Query Match 88.9%; Score 32; DB 21; Length 271;

Best Local Similarity 85.7%; Pred. No. 64;

Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GETRAPL 7

DB 79 GETRVLPL 85

RESULT 15

AAB60476

ID AAB60476 standard; Protein; 308 AA.

AC AAB60476;

DT 24-APR-2001 (first entry)

XX Human cell cycle and proliferation protein CCYPR-24, SEQ ID NO:24.

KW Cell cycle and proliferation protein; CCYPR; human; agonist;  
 KW antagonist; gene therapy; detection; gene therapy;  
 KW transgenic animal disease model; immune disorder;  
 KW developmental disorder; cell signalling disorder;  
 KW cell proliferative disorder; cancer; tumour; anaemia; epilepsy;  
 KW arteriosclerosis; asthma; allergy; diabetes mellitus;  
 KW menstrual cycle disorder; bacterial infection.

OS Homo sapiens.

PN WO200107471-A2.

PD 01-FEB-2001.

PF 21-JUL-2000; 2000WO-US19948.

PR 21-JUL-1999; 99US-0145075.

PR 08-SEP-1999; 99US-0153129.

PR 10-NOV-1999; 99US-0164647.

PA (INCY-) INCYTE GENOMICS INC.

PI Hillman JL, Lal P, Tang YT, Yue H, Au-Young J, Bandman O,

PI Azimzai Y, Yang J, Lu DM, Baughn WK, Patterson C, Shah P,

DR MPI; 2001-112727/12.

DR N-PSDB; AAF59613.

XX Human cell cycle and proliferation proteins and polynucleotides are

PT used to treat, diagnose and prevent immune, developmental and cell

PT signalling disorders and cell proliferative disorders including cancer -

PS Claim 1; Page 131-132; 205pp; English.

XX Sequences AAB60453-AAB60506 represent 54 human cell cycle and  
 CC proliferation proteins (CCYPR), which are encoded by AAF59590-AAF59643.  
 CC CCYPR and agonists of CCYPR are used to treat diseases or conditions  
 CC associated with decreased expression of functional CCYPR, while CCYPR  
 CC antagonists are used to treat diseases or conditions associated with  
 CC overexpression of functional CCYPR. Monoclonal or polyclonal antibodies  
 CC to CCYPR may be used in enzyme-linked immunosorbent assays (ELISA) or  
 CC radioimmunoassays to detect CCYPR. CCYPR itself may be used to detect  
 CC compounds e.g., antibodies, oligonucleotides and proteins (receptors)  
 CC that specifically bind to CCYPR, and in drug screening methods to  
 CC identify compounds that modulate the activity of CCYPR. CCYPR  
 CC nucleotides can be used to generate transgenic animal models of human  
 CC disease, and can be used in gene therapy in target cells with genetic  
 CC abnormalities with respect to the expression of CCYPR for the  
 CC treatment or prevention of a disorder associated with CCYPR.  
 CC Diseases which can be diagnosed, treated and prevented using CCYPR  
 CC proteins, nucleic acids, agonists or antagonists include immune,  
 CC developmental and cell signalling disorders, and cell proliferative  
 CC disorders including cancer. Specific examples of these disorders  
 CC include anaemia, epilepsy, arteriosclerosis, asthma, cancer, allergies,  
 CC diabetes mellitus, disorders of the menstrual cycle and infections  
 CC caused by bacteria.

SQ Sequence 308 AA;

Query Match 88.9%; Score 32; DB 22; Length 308;

Best Local Similarity 100.0%; Pred. No. 74;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GETRAP 6

DB 25 GETRAP 30

Search completed: January 29, 2003, 13:59:50  
 Job time : 27.3077 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2003, 13:57:01 ; Search time 9.87179 Seconds  
(without alignments)  
68.168 Million cell updates/sec

Title: US-09-807-949A-107

Perfect score: 36

Sequence: 1 GETRAPL 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	36	100.0	300	2 AB3188	acetyltransferase
2	32	88.9	412	2 UC4921	double C2 protein
3	31	86.1	273	2 E70939	hypothetical prote
4	31	86.1	1223	2 S29717	adenylate cyclase
5	30	83.3	84	2 AD2824	conserved hypotet
6	30	83.3	106	2 S51046	hypothetical prote
7	30	83.3	128	2 C97602	yeam protein (impo
8	30	83.3	168	2 T34969	hypothetical prote
9	30	83.3	344	2 AF3576	iron(III) diclrat
10	30	83.3	405	2 B82643	glucose dehydrogen
11	30	83.3	413	2 AH2933	N-carbamoyl-beta-a
12	30	83.3	441	2 G98348	n-carbamoyl-beta-a
13	30	83.3	479	2 F86285	F9L1.11 protein -
14	30	83.3	480	2 E86285	hypothetical prote
15	30	83.3	481	2 E86285	hypothetical prote
16	30	83.3	487	2 D86285	hypothetical prote
17	30	83.3	550	2 T34220	hypothetical prote
18	30	83.3	865	1 ISECTP	DNA topoisomerase
19	30	83.3	865	2 F90859	DNA topoisomerase
20	30	83.3	865	2 AB5760	DNA topoisomerase
21	30	83.3	865	2 AF0654	DNA topoisomerase
22	30	83.3	868	2 D83369	DNA topoisomerase
23	30	83.3	871	2 AB0370	DNA topoisomerase
24	30	83.3	876	2 B82163	DNA topoisomerase
25	30	83.3	1806	2 AF1717	probable peptidogl
26	30	83.3	1415	2 A46105	conserved hypotet
27	29	80.6	122	2 AI2992	NADH dehydrogenas
28	29	80.6	125	2 S53086	hypothetical prote
29	29	80.6	155	2 T15507	

30	29	80.6	218	2 A96291	hypothetical prote
31	29	80.6	274	2 H70003	probable prolyl am
32	29	80.6	361	2 A87383	flagellar biosynth
33	29	80.6	381	2 T36150	probable 4-hydroxy
34	29	80.6	408	2 B87436	conserved hypotet
35	29	80.6	416	2 G64542	glycine hydroxymet
36	29	80.6	421	2 E72573	probable 3-hydroxy
37	29	80.6	437	2 J02015	glucose dehydrogen
38	29	80.6	437	2 J02015	nucleoprotein - ra
39	29	80.6	437	2 J02016	nucleoprotein - ra
40	29	80.6	450	1 VHVNRV	nucleoprotein - ra
41	29	80.6	450	1 VHVNSB	nucleoprotein - ra
42	29	80.6	450	1 VHVNAV	nucleoprotein - ra
43	29	80.6	450	2 A46104	nucleoprotein N -
44	29	80.6	562	1 C75130	methylmalonyl-CoA
45	29	80.6	563	1 B71001	methylmalonyl-CoA

#### ALIGNMENTS

RESULT 1  
AB3188  
acetyltransferase [imported] - Agrobacterium tumefaciens (strain C58, Dupont) plasmid AT  
C:Species: Agrobacterium tumefaciens  
C:Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 11-Jan-2002  
C:Accession: AB3188  
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Marks, D.; Chen, L.; Wood, G.F.; Chen, Y.; Moo, L.  
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, W.; McEllell  
; Karp, P.; Romero, P.; Zhang, S.  
Science 294, 2317-2323, 2001  
A:Authors: Yoo, H.; Teo, Y.; Biddle, P.; Jung, M.; Kreppan, W.; Perry, M.; Gordon-Kamm, E  
ster, E.W.  
A>Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
A:Reference number: AB2577; PMID:11743193  
A:Accession: AB3188  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-300 <KIR>  
A:Cross-references: GB:AE008687; PIDN:AA145920.1; PID:g17743667; GSPDB:GN00188  
A:Experimental source: strain C58 (Dupont)  
C:Genetics:  
A:Gene: Atus231  
A:Genome: plasmid  
Query Match 100.0%; Score 36; DB 2; Length 300;  
Best local Similarity 100.0%; Pred. NO. 1.8;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GETRAPL 7  
DB 284 GETRAPL 290  
RESULT 2  
UC4921  
double C2 protein beta - mouse  
N:Alternate names: Doc2beta  
C:Species: Mus musculus (house mouse)  
C:Date: 26-Sep-1996 #sequence\_revision 01-Nov-1996 #text\_change 20-Jun-2000  
C:Accession: UC4921  
R:Kojima, T.; Fukuda, M.; Aruga, J.; Mikoshiba, K.  
U. Biochem. 120, 671-676, 1996  
A>Title: Calcium-dependent phospholipid binding to the C2A domain of a ubiquitous form of  
A:Reference number: UC4921; PMID:97058315; PMID:8902635  
A:Accession: UC4921  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-412 <KOJ>  
A:Cross-references: DDBJ:D85037; NID:g1565190; PIDN:BA12714.1; PID:g1565191  
C:Comment: This protein and rabphilin 3A are synaptic vesicle-associated proteins, and as  
C:Superfamily: synaptobrevin; protein kinase C C2 region homology  
F;120-236/Domain: protein kinase C C2 region homology <KC2A>

F:123-236/Domain: C2A #status predicted <C2A>  
F:260-375/Domain: protein kinase C C2 region homology <KC2B>  
F:264-375/Domain: C2B #status predicted <C2B>

Query Match 88.9%; Score 32; DB 2; Length 412;  
Best Local Similarity 85.7%; Pred. No. 21;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GETRAPL 7  
DB 229 GETRVPL 235

## RESULT 3

E70939  
hypothetical protein RV0249c - Mycobacterium tuberculosis (strain H37RV)

C/Species: Mycobacterium tuberculosis  
C/Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000  
C/Accession: E70939

R/Colo, S.T.; Broesch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Felwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, R.; Rutter, S.; Seeger, K.; Skellom, S.; Squares, S.  
Nature 393, 537-544, 1998

A/Authors: Squares, R.; Sulistion, J.E.; Taylor, K.; Whitehead, S.; Barrett, B.G.  
A/Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A/Reference number: A70500; MUID:98295987; PMID:9634230

A/Accession: E70939

A/Status: preliminary; nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-273 <COL>

A/Cross-references: GB:AL021929; GB:AL123456; NID:g3242291; PIDN:CAA17341.1; PID:g290945

A/Experimental source: strain H37RV

C/Genetics:

A/Gene: RV0249c

C/Superfamily: Mycobacterium tuberculosis hypothetical protein RV0249c

Query Match 86.1%; Score 31; DB 2; Length 273;  
Best Local Similarity 85.7%; Pred. No. 23;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GETRAPL 7  
DB 143 GETRLPL 149

## RESULT 4

S29717  
adenylate cyclase (EC 4.6.1.1) type 5 - rat

C/Species: Rattus norvegicus (Norway rat)  
C/Date: 25-Feb-1994 #sequence\_revision 01-Sep-1995 #text\_change 04-Sep-1998  
C/Accession: S29717; B47202

R/Glaty, C.E.; Snyder, S.H.

Nature 361, 536-538, 1993

A/Title: Cloning and expression of an adenylyl cyclase localized to the corpus striatum.  
A/Reference number: S29717; MUID:93156838; PMID:8429907

A/Accession: S29717

A/Status: preliminary

A/Molecule type: mRNA

A/Residues: 1-1223 <GLA>

R/Premont, R.T.; Chen, J.; Ma, H.W.; Ponnappalli, M.; Iyengar, R.  
Proc. Natl. Acad. Sci. U.S.A. 89, 9809-9813, 1992

A/Title: Two members of a widely expressed subfamily of hormone-stimulated adenylyl cyclase  
A/Reference number: A47202; MUID:93028552; PMID:1403703

A/Accession: B47202

A/Status: preliminary; not compared with conceptual translation

A/Molecule type: mRNA

A/Residues: 91,'GKG','95-97','E','134-183','R','185-209','V','211-237','A','239-346','I','348-511','RE>

A/Experimental source: liver, kidney

A/Note: sequence extracted from NCBI backbone (NCBI:115852)

C/Superfamily: human adenylyl cyclase; guanylate cyclase catalytic domain homology

C/Keywords: phosphorus-oxygen lyase

F:373-608/Domain: guanylate cyclase catalytic domain homology <GCC>

F:981-1220/Domain: guanylate cyclase catalytic domain homology <GCC2>

Query Match 86.1%; Score 31; DB 2; Length 1223;  
Best Local Similarity 85.7%; Pred. No. 1.2e+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GETRAPL 7  
DB 88 GETRAPL 94

## RESULT 5

AD2824

conserved hypothetical protein Atu2018 [imported] - Agrobacterium tumefaciens (strain C58)

C/Species: Agrobacterium tumefaciens  
C/Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 11-Jan-2002

C/Accession: AD2824

R/Wood, D.W.; Setubal, J.C.; Kaul, R.; Monke, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.  
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McCellie,  
Karp, P.; Romero, P.; Zhang, S.  
Science 294, 2317-2323, 2001

A/Authors: Yoo, H.; Tao, Y.; Biddle, P.; Tung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,  
ster, E.W.

A/Title: The genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
A/Reference number: AB2577; PMID:11743193

A/Accession: AD2824

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-84 <GB>

A/Cross-references: GB:AE008688; PIDN:AAL43010.1; PID:g17740473; GSPDB:GN00186

A/Experimental source: strain C58 (Dupont)

C/Genetics:

A/Gene: Atu2018

A/Map position: circular chromosome

Query Match 83.3%; Score 30; DB 2; Length 84;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 ETRAPL 7  
DB 22 ETRAPL 27

## RESULT 6

S51046  
hypothetical protein 1 - Paracoccus denitrificans

C/Species: Paracoccus denitrificans  
C/Date: 01-Aug-1995 #sequence\_revision 01-Sep-1995 #text\_change 08-Oct-1999  
C/Accession: S51046

R/van Spanning, R.J.M.; van der Pelen, C.J.N.M.; Slotboom, D.J.; Reijnders, W.N.M.; Stout  
Eur. J. Biochem. 226, 201-210, 1994

A/Title: Expression of the mau genes involved in methylamine metabolism in Paracoccus der  
A/Reference number: S51046; MUID:95045590; PMID:757249

A/Accession: S51046

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-106 <VAN>

A/Cross-references: EMBL:U12464; NID:g558801; PIDN:AA56721.1; PID:g558802

Query Match 83.3%; Score 30; DB 2; Length 106;  
Best Local Similarity 71.4%; Pred. No. 14;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 GETRAPL 7  
DB 42 GQTRAPV 48

## RESULT 7

C97602  
yefm protein [imported] - Agrobacterium tumefaciens (strain C58, Cereon)

C/Species: Agrobacterium tumefaciens

C>Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 11-Jan-2002  
 C/Accession: C97602  
 R/Gooner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Ourullo, B.; Goldman, A.; Liu, F.; Molinar, C.; Allinger, M.; Dougherty, D.; Scott, C.; Lappas, C.; Markels, B.; Science 294, 2323-2328, 2001  
 A/Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent *Agrobacterium tumefaciens*  
 A/Reference number: A97359; PMID:11743194  
 A/Accession: C97602  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-128 <KUR>  
 A/Cross-references: GB:AE007869; PIDN:AAK87772.1; PID:g15157142; GSPDB:GN00169  
 C/Genetics:  
 A/Map position: circular chromosome

Query Match 83.3%; Score 30; DB 2; Length 128;  
 Best Local Similarity 100.0%; Pred. No. 17;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GETRAPL 7  
 |||||  
 Db 66 GETRAPL 71

RESULT 8  
 T34969  
 hypothetical protein SC4A10.32 - Streptomyces coelicolor  
 C/Species: Streptomyces coelicolor  
 C/Date: 05-Nov-1999 #sequence\_revision 05-Nov-1999 #text\_change 05-Nov-1999  
 R/Saunders, D.C.; Harris, D.; James, K.D.; Parthill, J.; Barrett, B.G.; Rajandream, M.A. submitted to the EMBL Data Library, August 1999  
 A/Reference number: Z21563  
 A/Accession: T34969  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: DNA  
 A/Residues: 1-168 <SAU>  
 A/Cross-references: EMBL:AL09663; PIDN:CA852008.1; GSPDB:GN00070; SCQEDB:SC4A10.32  
 A/Experimental source: strain A3(2)  
 C/Genetics:  
 A/Map position: SC4A10.32

Query Match 83.3%; Score 30; DB 2; Length 168;  
 Best Local Similarity 85.7%; Pred. No. 23;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GETRAPL 7  
 |||||  
 Db 71 GETRAPL 77

RESULT 9  
 AF3576  
 Iron(III) dictrate-binding periplasmic protein BME110535 [imported] - *Brucella melitensis*  
 C/Species: *Brucella melitensis*  
 C/Date: 01-Feb-2002 #sequence\_revision 01-Feb-2002 #text\_change 01-Feb-2002  
 R/Accession: AF3576  
 R/DeVos, V.G.; Kapata, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Lo, T.; Ivanova, P.; Mazur, M.; Goldman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letens, Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002  
 A/Title: The genome sequence of the facultative intracellular pathogen *Brucella melitensis*  
 A/Reference number: AD3252; PMID:11756688  
 A/Accession: AF3576  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-344 <KUR>  
 A/Cross-references: GB:AE008918; PIDN:AAL53777.1; PID:g17984706; GSPDB:GN00191  
 A/Experimental source: strain 16M  
 C/Genetics:  
 A/Map position: II

Query Match 83.3%; Score 30; DB 2; Length 344;  
 Best Local Similarity 85.7%; Pred. No. 51;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GETRAPL 7  
 |||||  
 Db 30 GETRAPL 36

RESULT 10  
 B82643  
 glucose dehydrogenase B XP1740 [imported] - *Xylella fastidiosa* (strain 9a5c)  
 C/Species: *Xylella fastidiosa*  
 C/Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Sep-2000  
 R/Accession: B82643  
 R/anonymous, The *Xylella fastidiosa* Consortium of the Organization for Nucleotide Sequencing  
 Nature 406, 151-157, 2000  
 A/Title: The genome sequence of the plant pathogen *Xylella fastidiosa*.  
 A/Reference number: A82515; PMID:20365717; PMID:10910347  
 A/Note: for a complete list of authors see reference number A59328 below  
 A/Accession: B82643  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-405 <SIM>  
 A/Cross-references: GB:AE003997; GB:AE003849; NID:g9106805; PIDN:AAF4549.1; GSPDB:GN0012

A/Experimental source: strain 9a5c  
 R/Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvares, R.; A. Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carreir, H. as-Neto, E.; Docena, C.; El-Dorri, H.; Facchini, A.P.; Ferreira, A.V.S. submitted to GenBank, June 2000  
 A/Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm, J.D.; Junqueira, M.L.; Kemper, E.L.; Kltajima, J.P.; Krieger, J.B.; Kuramae, E.E.; Laigre, Chdo, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E.

A/Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.; F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A.; Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santeilli, R.V.; Sawaaki, A.; Authours: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveira, M.; Tshahko, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Zé  
 A/Reference number: A59328  
 A/Contents: annotation  
 C/Genetics:  
 A/Map position: XP1740

C/Superfamily: hypothetical protein b0837

Query Match 83.3%; Score 30; DB 2; Length 405;  
 Best Local Similarity 71.4%; Pred. No. 60;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 GETRAPL 7  
 |||||  
 Db 49 GETRAPL 55

RESULT 11  
 AH2933  
 N-Carboxyl-beta-alanine amidohydrolase [imported] - *Agrobacterium tumefaciens* (strain C)  
 C/Species: *Agrobacterium tumefaciens*  
 C/Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 01-Feb-2002  
 R/Accession: AH2933  
 R/Wood, D.W.; Seubul, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.; erge, G.; Giller, W.; Grant, C.; Guenther, D.; Kuyavlin, T.; Levy, R.; Li, M.; McCellie  
 Science 294, 2317-2323, 2001  
 A/Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, E. ster, E.W.  
 A/Title: The Genome of the Natural Genetic Engineer *Agrobacterium tumefaciens* C58.  
 A/Reference number: AB2577; PMID:11743193  
 A/Accession: AH2933  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-413 <KUR>  
 A/Cross-references: GB:AE008689; PIDN:AAL43886.1; PID:g17741433; GSPDB:GN00187  
 A/Experimental source: strain C58 (Dupont)

C:Genetics:  
A:Gene: amab  
A:Map position: linear chromosome  
C:Superfamily: N-carbamyl-L-amino acid amidohydrolyase

Query Match 83.3%; Score 30; DB 2; Length 413;  
Best Local Similarity 100.0%; Pred. No. 62;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 73 ETRAPL 78

RESULT 12  
G98348  
N-carbamoyl-beta-alanine amidohydrolyase PA0444 [imported] - Agrobacterium tumefaciens (s  
C:Species: Agrobacterium tumefaciens  
C>Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 11-Jan-2002  
C:Accession: G98348  
R:Goodner, B.; Hinkle, G.; Galtung, S.; Miller, N.; Blanchard, M.; Quirillo, B.; Goldman,  
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;  
Science 294, 2323-2328, 2001  
A>Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tun  
A:Reference number: A97359; PMID:11743194  
A:Accession: G98348  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-441 <KOR>  
A:Cross-references: GB:AE007870; PIDN:AAK90313.1; PID:G51560344; GSPDB:GN00170  
C:Genetics:  
A:Gene: AGR\_L\_3478  
A:Map position: linear chromosome  
C:Superfamily: N-carbamyl-L-amino acid amidohydrolyase

Query Match 83.3%; Score 30; DB 2; Length 441;  
Best Local Similarity 100.0%; Pred. No. 66;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 101 ETRAPL 106

RESULT 13  
F86285  
F9L11 protein - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C>Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 02-Nov-2001  
C:Accession: F86285  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Creasy, T.H.; Dewar, K.;  
ansen, N.F.; Hughes, B.; Huizart, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziani,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shim, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A>Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719; PMID:11130712  
A:Accession: F86285  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-479 <STO>  
A:Cross-references: GB:AE005172; NID:G5103814; PIDN:AAD39644.1; GSPDB:GN00141  
C:Genetics:  
A:Map position: 1

Query Match 83.3%; Score 30; DB 2; Length 479;  
Best Local Similarity 100.0%; Pred. No. 73;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 2 ETRAPL 7  
Db 243 ETRAPL 248

RESULT 14  
G86285  
hypothetical protein F9L12 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C>Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Dec-2001  
C:Accession: G86285  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Creasy, T.H.; Dewar, K.;  
ansen, N.F.; Hughes, B.; Huizart, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziani,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shim, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A>Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719; PMID:11130712  
A:Accession: G86285  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-480 <STO>  
A:Cross-references: GB:AE005172; NID:G5103816; PIDN:AAD39646.1; GSPDB:GN00141  
C:Genetics:  
A:Map position: 1

Query Match 83.3%; Score 30; DB 2; Length 480;  
Best Local Similarity 100.0%; Pred. No. 73;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 244 ETRAPL 249

RESULT 15  
E86285  
hypothetical protein F9L10 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C>Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Dec-2001  
C:Accession: E86285  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Creasy, T.H.; Dewar, K.;  
ansen, N.F.; Hughes, B.; Huizart, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziani,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shim, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A>Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719; PMID:11130712  
A:Accession: E86285  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-481 <STO>  
A:Cross-references: GB:AE005172; NID:G5103818; PIDN:AAD39648.1; GSPDB:GN00141  
C:Genetics:  
A:Map position: 1

Query Match 83.3%; Score 30; DB 2; Length 481;  
Best Local Similarity 100.0%; Pred. No. 73;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Cy 2 ETRAPL 7  
Db 240 ETRAPL 245

Thu Jan 30 11:47:37 2003

us-09-807-949a-107.rpr

Page 5

Search completed: January 29, 2003, 14:03:22  
Job time : 10.8718 secs

---

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2003, 13:51:40 ; Search time 4.84615 Seconds  
(without alignments)  
59,910 Million cell updates/sec

Title: US-09-807-949A-107

Perfect score: 36

Sequence: 1 GETRAPL 7

Scoring table: BIOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	32	88.9	308	1 MGPI_HUMAN	Q9h4y2 homo sapien
2	32	88.9	887	1 ORP3_HUMAN	Q9h415 homo sapien
3	30	83.3	865	1 TOP1_BCOLI	P06612 escherichia
4	30	83.3	865	1 TOP1_SALTY	P40686 salmonella
5	30	83.3	868	1 TOP1_PSEAE	Q9h315 pseudomonas
6	30	83.3	876	1 TOP1_VIBCH	Q9h3d2 vibrio chol
7	30	83.3	3415	1 POLG_POWVL	Q04538 t genome po
8	29	80.6	274	1 YTYM_BACST	P23574 bacillus su
9	29	80.6	416	1 GLYA_HELPY	P56089 helicobacte
10	29	80.6	421	1 HMDH_AERPE	Q9y864 aeropyrum p
11	29	80.6	450	1 NCAP_PABVA	P15197 rabies viru
12	29	80.6	450	1 NCAP_PABVF	Q08314 rabies viru
13	29	80.6	450	1 NCAP_PABVP	P06025 rabies viru
14	29	80.6	450	1 NCAP_PABVS	P16285 rabies viru
15	29	80.6	450	1 NCAP_PABVU	Q09110 rabies viru
16	29	80.6	542	1 MODU_DROME	P13469 drosophila
17	29	80.6	936	1 MSH4_HUMAN	Q15457 homo sapien
18	29	80.6	1300	1 SAL3_HUMAN	Q9bxa9 homo sapien
19	28	77.8	92	1 Y676_TREPA	O83682 treponema p
20	28	77.8	201	1 Y048_MYCTU	P21705 mycobacteri
21	28	77.8	289	1 Y048_MYCTU	Q06080 vibrio mini
22	28	77.8	316	1 DHAS_VIBMT	P23247 vibrio chol
23	28	77.8	337	1 DHAS_VIBCH	Q9zces listeria in
24	28	77.8	415	1 PROA_LISIN	Q9zces listeria in
25	28	77.8	415	1 PROA_LISMO	Q93455 listeria mo
26	28	77.8	417	1 YEIM_HABIN	P44742 haemophilus
27	28	77.8	464	1 CAPB_BACAN	P19580 bacillus an
28	28	77.8	631	1 DNAA_PASHA	O52064 pasteurella
29	28	77.8	631	1 DNAA_THERO	P96133 thermotacro
30	28	77.8	632	1 DNAA_ACTAC	P71331 actinobacill
31	28	77.8	633	1 DNAA_HAEDU	P48209 haemophilus
32	28	77.8	634	1 DNAA_HABIN	P43736 haemophilus
33	28	77.8	634	1 DNAA_PASWU	P57870 pasteurella

34	28	77.8	637	1 DNAA_PSEAE	Q9h4y3 pseudomonas
35	28	77.8	638	1 DNAA_PSEEG	Q9hwg9 pseudomonas
36	28	77.8	642	1 DNAA_FRATU	P48205 francisella
37	28	77.8	644	1 DNAA_LEGPN	O32482 legionella
38	28	77.8	933	1 PERT_HUMAN	P07202 homo sapien
39	28	77.8	938	1 YMG1_CAEBL	Q22712 caenorhabdi
40	28	77.8	1333	1 VGR1_MOUSE	P35969 mus muscula
41	28	77.8	1336	1 VGR1_RAT	P53767 rattus norv
42	28	77.8	1338	1 VGR1_HUMAN	P17948 h vascular
43	27	75.0	133	1 PRTL_ECOLI	P02338 escherichia
44	27	75.0	133	1 RSG_SYNY3	P73307 synecocyst
45	27	75.0	143	1 RKZ_SOYBN	P18663 glycine max

## ALIGNMENTS

```

RESULT 1
MGPI_HUMAN          STANDARD;          PRT;          308 AA.
ID MGPI_HUMAN
AC Q9H4Y2; Q9H215;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Melanoma-associated antigen F1 (MAGE-F1 antigen).
GN MAGEF1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Store B.C., Schummer M., Paley P.J., Crawford M., Ford M.,
RA Nelson B.H.;
RT "MAGE-F1, a novel ubiquitously expressed member of the MAGE
RT superfamily identified by SEREX immunoscreening."
RT Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-284 FROM N.A.
RC TISSUE=Kidney;
RA Lucas S., Boon T.;
RT "Identification of new genes of the MAGE family."
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
CC -1- TISSUE SPECIFICITY: Ubiquitous.
CC -1- SIMILARITY: CONTAINS 1 MAGE DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-eb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF295378; AAG30208.1; -
CC EMBL; AF320910; AAG38606.1; -
CC InterPro: IPR002190; MAGE.
CC Pfam: PF01454; MAGE; 1.
CC PROSITE: PS50838; MAGE; 1.
CC
CC DOMAIN
CC
CC SEQUENCE 308 AA; 35308 MW; 8FB0092653090580 CRC64;
CC
CC
CC Query Match 88.9%; Score 32; DB 1; Length 308;
CC Best Local Similarity 100.0%; Pred. No. 6.5;
CC Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC
CC QY 1 GETRAP 6
CC
CC DB 25 GETRAP 30
CC
CC RESULT 2

```

ORP3\_HUMAN STANDARD; PRT; 887 AA.  
 AC OSH45; O9BZF4; O14591; OSUEDS; O43357; O43358;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Oxyesterol binding protein-related protein 3 (OSBP-related protein 3)  
 DE (ORP-3).  
 GN OSBP13 OR ORP3 OR KIAA0704.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RA Gregorio-King C.C., Collier G.C., McMillan J., Waugh C.,  
 RA Kirkland M.A.;  
 RT "ORP3: a novel human oxysterol-binding protein gene expressed in CD34+  
 RT stem/progenitor cells";  
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.  
 [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21601154; PubMed=11735225;  
 RA Jaworski C.J., Moreira E., Li A., Lee R., Rodriguez I.R.;  
 RT "A family of 12 human genes containing oxysterol-binding domains.";  
 RL Genomics 78:185-196(2001).  
 [3]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Uterus;  
 RA Strausberg R.;  
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.  
 [4]  
 RP SEQUENCE OF 1-245 FROM N.A.  
 RX MEDLINE=21376257; PubMed=11483621;  
 RA Lehto M., Laitinen S., Chinetti G., Johansson M., Ehnholm C.,  
 RA Saele B., Ikonen E., Oikonen V.M.;  
 RT "The OSBP-related protein family in humans.";  
 RL J. Lipid Res. 42:1203-1213(2001).  
 [5]  
 RP SEQUENCE OF 221-887 FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=98403880; PubMed=9734811;  
 RA Ishikawa K.-I., Nagase T., Suyama M., Miyajima N., Tanaka A.,  
 RA Kotani H., Nomura N., Ohara O.;  
 RT "Prediction of the coding sequences of unidentified human genes. X.  
 RT The complete sequences of 100 new cDNA clones from brain which can  
 RT code for large proteins in vitro.";  
 RL DNA Res. 5:169-176(1998).  
 [6]  
 RP SEQUENCE OF 291-467 FROM N.A.  
 RA Dante M., Kellen J., O'Brian D.;  
 RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.  
 [7]  
 RP SEQUENCE OF 500-649 FROM N.A.  
 RA Dante M., Kellen J., O'Brian D.;  
 RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.  
 [8]  
 RP SEQUENCE OF 676-887 FROM N.A.  
 RA Dauphin S., Mooney A.;  
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: BELONGS TO THE OSBP FAMILY.  
 CC -1- SIMILARITY: CONTAINS 1 PH DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 CC DR EMBL; AY008372; AAC23400.1; -;  
 CC EMBL; AF392444; AAL40657.1; -;  
 DR

DR EMBL; BC017731; AAH17731.1; -;  
 DR EMBL; AB014604; BAA31679.1; -;  
 DR EMBL; AF323727; AAG53408.1; -;  
 DR EMBL; AC003093; AAB83939.1; -;  
 DR EMBL; AC004016; AAC26986.1; -;  
 DR EMBL; AC004016; AAC26985.1; -;  
 DR Genew; HGNC:16370; OSBP13.  
 DR MIM; 606732; -;  
 DR InterPro; IPR006048; Oxysterol\_BP.  
 DR InterPro; IPR001849; PH.  
 DR Pfam; PF00169; PH; 1.  
 DR Pfam; PF01237; Oxysterol\_BP; 1.  
 DR SMART; SMO0233; PH; 1.  
 DR PROSITE; PS01013; OSBP; 1.  
 DR PROSITE; PSS0003; PH DOMAIN; 1.  
 KW Lipid transport; Transport.  
 FT DOMAIN 51 146 PH.  
 FT CONFLICT 676 676 V -> S (IN REF. 8).  
 SQ SEQUENCE 887 AA; 101223 MW; 12E16912BD3F2E99 CRC64;  
 Query Match 88.9%; Score 32; DB 1; Length 887;  
 Best Local Similarity 85.7%; Pred. No. 21;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 Qy 1 GETRAPL 7  
 Db 205 GETRAPL 211  
 ID TOP1\_ECOLI STANDARD; PRT; 865 AA.  
 AC P06612;  
 DT 01-JAN-1988 (Rel. 06, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE DNA topoisomerase I (EC 5.99.1.2) (Omega-protein) (Relaxing enzyme)  
 DE (untwisting enzyme) (Swivelase).  
 GN TOPA OR SUPX OR B1274.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=87141163; PubMed=3029379;  
 RA Tse-Dinh Y.-C., Wang J.C.;  
 RT "Complete nucleotide sequence of the topA gene encoding Escherichia  
 RT coli DNA topoisomerase I.";  
 RL J. Mol. Biol. 191:321-331(1986).  
 [2]  
 RP REVISIONS, SEQUENCE FROM N.A.  
 RA Lynch D.A., Wang J.C.;  
 RL Submitted (SEP-1991) to the EMBL/GenBank/DBJ databases.  
 [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / M61655;  
 RX MEDLINE=97426617; PubMed=9278503;  
 RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-Vides J., Glaesner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.;  
 RT "The complete genome sequence of Escherichia coli K-12.";  
 RL Science 277:1245-1247(1997).  
 [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=97251357; PubMed=9097039;  
 RA Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K.,  
 RA Itoh T., Kasai H., Kashimoto K., Kimura S., Kitakawa M.,  
 RA Kitagawa M., Makino K., Miki T., Mizobuchi K., Mori H., Mori T.,  
 RA Motomura K., Nakade S., Nakamura Y., Nashimoto H., Nishio Y.,  
 RA Ohshima T., Saito N., Sampei G., Seki Y., Sivasubram S.,



RA Tagami H., Takeda J., Takemoto K., Takeuchi Y., Wada C.,  
 RA Yamamoto Y., Horiuchi T.;  
 RT "A 570-bp DNA sequence of the *Escherichia coli* K-12 genome  
 RT corresponding to the 28.0-40.1 min region on the linkage map.";  
 RL DNA Res. 3:363-377(1996).  
 RN [5]  
 RP SEQUENCE OF 696-865 FROM N.A.  
 RX MEDLINE=87194810; PubMed=3032952;  
 RA Ostrowski J., Jagura-Burdzy G., Kredich N.M.;  
 RT "DNA sequences of the *cyoB* regions of *Salmonella typhimurium* and  
 RT *Escherichia coli*.";  
 RL J. Biol. Chem. 262:5999-6005(1987).  
 RN [6]  
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1-590.  
 RX MEDLINE=94159070; PubMed=8114910;  
 RA Lima C.D., Wang J.C., Mondragon A.;  
 RT "Three-dimensional structure of the 67K N-terminal fragment of E.  
 RT *coli* DNA topoisomerase I.";  
 RL Nature 367:138-146(1994).  
 RN [7]  
 RP STRUCTURE BY NMR OF 745-865.  
 RX MEDLINE=95298771; PubMed=7779808;  
 RA Yu L., Zhu C.-X., Tse-Dinh Y.-C., Festk S.W.;  
 RT "Solution structure of the C-terminal single-stranded DNA-binding  
 RT domain of *Escherichia coli* topoisomerase I.";  
 RL Biochemistry 34:7622-7628(1995).  
 CC -1- FUNCTION: THE REACTION CATALYZED BY TOPOISOMERASES LEADS TO THE  
 CC CONVERSION OF ONE TOPOLOGICAL ISOMER OF DNA TO ANOTHER.  
 CC -1- CATALYTIC ACTIVITY: ATP-independent breakage of single-stranded  
 CC DNA, followed by passage and rejoining.  
 CC -1- SUBUNIT: MONOMER.  
 CC -1- MISCELLANEOUS: WHEN A TOPOISOMERASE TRANSIENTLY BREAKS A DNA  
 CC BACKBONE BOND, IT SIMULTANEOUSLY FORMS A PROTEIN-DNA LINK, IN  
 CC WHICH A TYROSYL OXYGEN IN THE ENZYME IS JOINED TO A DNA PHOSPHORUS  
 CC AT ONE END OF THE ENZYME-SEVERED DNA STRAND.  
 CC -1- SIMILARITY: BELONGS TO PROKARYOTIC TYPE I/III TOPOISOMERASE  
 CC FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sb-sib.ch](mailto:license@sb-sib.ch)).  
 CC -----  
 CC EMBL; X04475; CA28164.1; -;  
 CC EMBL; M15041; AA23641.1; -;  
 CC EMBL; AE000225; AAC74356.1; -;  
 CC EMBL; D90764; BA14811.1; -;  
 CC EMBL; D90765; BA14826.1; -;  
 CC PIR; A25786; ISECTP.  
 CC PDB; 1ECU; 31-JUL-95.  
 CC PDB; 1YUA; 08-MAR-96.  
 CC EC02DBAB; 1115.0; 6TH EDITION.  
 CC EcoGene; EG11013; topA.  
 CC InterPro; IPR002936; DNAPrim\_toprim.  
 CC InterPro; IPR003601; DNATopI\_ATP\_bind.  
 CC InterPro; IPR003602; DNATopI\_DNA\_bind.  
 CC InterPro; IPR000380; Prok\_topisomerase.  
 CC Pfam; PF01131; Topoisom Dac; 1.  
 CC Pfam; PF01396; zf-C4\_Topoisom; 2.  
 CC Pfam; PF01751; Toprim; 1.  
 CC PRINTS; PRO0417; PRTPISMRASE1.  
 CC SMART; SM00437; TOP1AC; 1.  
 CC SMART; SM00436; TOP1BC; 1.  
 CC SMART; SM00493; TOPRIM; 1.  
 CC TIGRFAMs; TIGR01051; topA\_bact; 1.  
 CC PROSITE; PS00396; TOPOISOMERASE\_I\_PROK; 1.  
 CC IsoMeras; Topoisomerase; DNA-binding; Zinc-finger; Metal-binding;  
 KW Repeat; 3d-structure; Complete proteome.  
 FT ZN\_FING 599 630 C4-TYPE 1.  
 FT ZN\_FING 662 689 C4-TYPE 2.

FT ZN\_FING 711 736 C4-TYPE 3.  
 FT ACT\_SITE 319 319 DNA\_CLEAVAGE.  
 FT CONFLICT 787 787 P -> R (IN REF. 5).  
 SO SEQUENCE 865 AA; 97349 MW; 8C13F767F5B178C CRC64;  
 Query Match 83.3%; Score 30; DB 1; Length 865;  
 Best Local Similarity 100.0%; Pred. No. 59;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 ETRAPL 7  
 DB 791 ETRAPL 796  
 RESULT 4  
 ID TOP1\_SALTLY STANDARD; PRT; 865 AA.  
 AC P40686;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 15-FEB-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE DNA topoisomerase I (EC 5.99.1.2) (Omega-protein) (Relaxing enzyme)  
 DE (Unwinding enzyme) (Swivelase).  
 GN TOPA OR STM1714 OR STY1336.  
 OS *Salmonella typhimurium*, and  
 OS *Salmonella typhi*.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC *Salmonella*.  
 OX NCBI\_Taxid=602, 601;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=S.typhimurium; STRAIN=L72 / SGSC1412 / ATCC 700720;  
 RX MEDLINE=21534948; PubMed=11675609;  
 RA McClelland M., Sanders K.E., Speeth J., Clifton S.W., Lettelle P.,  
 RA Courtney L., Porwollik S., Ali D., Dante M., Du F., Hou S., Layman D.,  
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,  
 RA Ryan E., Sun H., Flores L., Miller W., Stoneking T., Nhan M.,  
 RA Waterston R., Wilson R.K.;  
 RT "Complete genome sequence of *Salmonella enterica* serovar Typhimurium  
 RT L72.";  
 RL Nature 413:852-856(2001).  
 RN [2]  
 RP SEQUENCE OF 856-865 FROM N.A.  
 RC SPECIES=S.typhimurium;  
 RX MEDLINE=87194810; PubMed=3032952;  
 RA Ostrowski J., Jagura-Burdzy G., Kredich N.M.;  
 RT "DNA sequences of the *cyoB* regions of *Salmonella typhimurium* and  
 RT *Escherichia coli*.";  
 RL J. Biol. Chem. 262:5999-6005(1987).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=S.typhi; STRAIN=CT18;  
 RX MEDLINE=21534947; PubMed=11677608;  
 RA Parikh J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,  
 RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebatista M.,  
 RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,  
 RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,  
 RA Felwell T., Hamlin N., Haque A., Hien T.T., Holtroyd S., Jagels K.,  
 RA Krogh A., Larsen T.S., Leather S., Moule S., O'Garra P., Parry C.,  
 RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,  
 RA Whitehead S., Barrrell B.G.;  
 RT "Complete genome sequence of a multiple drug resistant *Salmonella*  
 RT *enterica* serovar Typhi CT18.";  
 RL Nature 413:848-852(2001).  
 RN [4]  
 RP FUNCTION: THE REACTION CATALYZED BY TOPOISOMERASES LEADS TO THE  
 CC CONVERSION OF ONE TOPOLOGICAL ISOMER OF DNA TO ANOTHER.  
 CC -1- CATALYTIC ACTIVITY: ATP-independent breakage of single-stranded  
 CC DNA, followed by passage and rejoining.  
 CC -1- SUBUNIT: MONOMER (BY SIMILARITY).  
 CC -1- MISCELLANEOUS: WHEN A TOPOISOMERASE TRANSIENTLY BREAKS A DNA  
 CC BACKBONE BOND, IT SIMULTANEOUSLY FORMS A PROTEIN-DNA LINK, IN  
 CC WHICH A TYROSYL OXYGEN IN THE ENZYME IS JOINED TO A DNA PHOSPHORUS  
 CC AT ONE END OF THE ENZYME-SEVERED DNA STRAND.

CC -1- SIMILARITY: BELONGS TO PROKARYOTIC TYPE I/III TOPOISOMERASE  
 CC FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC DR EMBL; AEO08776; AAL20632.1; -  
 CC DR EMBL; M15040; AAA27044.1; -  
 CC DR EMBL; AL627269; CAD08417.1; -  
 CC DR StvGene; SG10515; topA.  
 CC DR InterPro; IPR002936; DNAPrim\_toprim.  
 CC DR InterPro; IPR00380; Prok\_Episomase.  
 CC DR Pfam; PF01131; Topoisom\_bac; 1.  
 CC DR Pfam; PF01396; zf-C4\_Topoiso; 2.  
 CC DR Pfam; PF01751; Toprim; 1.  
 CC DR TIGRFAMs; TIGR01051; topA\_bact; 1.  
 CC DR PROSITE; PS00396; TOPOISOMERASE\_I\_PROK; 1.  
 CC DR Isomerase; topoisomerase; DNA-binding; Zinc-finger; Metal-binding;  
 CC Repeat; Complete proteome.  
 CC FT ZN\_FING 599 630 C4-TYPE 1.  
 CC FT ZN\_FING 662 689 C4-TYPE 2.  
 CC FT ZN\_FING 711 736 C4-TYPE 3.  
 CC FT ACT\_SITE 319 319 DNA\_CLEAVAGE (BY SIMILARITY).  
 CC SQ SEQUENCE 865 AA; 97302 MW; 57BE1CBA696FEF7C CRC64;  
 CC  
 CC Query Match 83.3%; Score 30; DB 1; Length 865;  
 CC Best Local Similarity 100.0%; Pred. No. 59;  
 CC Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 CC  
 CC QY 2 ETRAPL 7  
 CC Db 791 ETRAPL 796  
 CC  
 CC RESULT 5  
 CC ID TOP1\_PSEAE STANDARD; PRT; 868 AA.  
 CC AC Q9H2J5;  
 CC DT 16-OCT-2001 (Rel. 40, Created)  
 CC DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 CC DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 CC DE DNA topoisomerase I (EC 5.99.1.2) (Omega-protein) (Relaxing enzyme)  
 CC DE (Unwisting enzyme) (Swivelase).  
 CC DE TOPA OR PA3011.  
 CC OS Pseudomonas aeruginosa.  
 CC OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
 CC OC Pseudomonas.  
 CC NCBI\_TaxID=287;  
 CC RN NCI [1]  
 CC RP SEQUENCE FROM N.A.  
 CC RX STRAIN=ATCC 15692 / PAOI;  
 CC MEDLINE=20437337; PubMed=10984043;  
 CC RA Hickey M.J., Brinkman F.S.L., Hutnagle W.O., Kowalik D.J., Lagrou M.,  
 CC Gader R.L., Goltzy L., Tolentino E., Westbrock-Medman S., Yuan Y.,  
 CC Brody L.L., Coulter S.N., Folger K.R., Kas A., Lardig K., Lam R.M.,  
 CC Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,  
 CC Reizer J., Slier M.H., Hancock R.E.W., Lory S., Olson M.V.;  
 CC "Complete genome sequence of Pseudomonas aeruginosa PAOI, an  
 CC opportunistic pathogen.";  
 CC Nature 406:959-964(2000).  
 CC RL Nature 406:959-964(2000).  
 CC CC -1- FUNCTION: THE REACTION CATALYZED BY TOPOISOMERASES LEADS TO THE  
 CC CONVERSION OF ONE TOPOLOGICAL ISOMER OF DNA TO ANOTHER.  
 CC CC -1- CATALYTIC ACTIVITY: ATP-independent breakage of single-stranded  
 CC DNA, followed by passage and rejoining.  
 CC CC -1- SUBUNIT: MONOMER (BY SIMILARITY).  
 CC CC -1- MISCELLANEOUS: WHEN A TOPOISOMERASE TRANSIENTLY BREAKS A DNA  
 CC BACKBONE BOND, IT SIMULTANEOUSLY FORMS A PROTEIN-DNA LINK, IN

CC WHICH A TYROSYL OXYGEN IN THE ENZYME IS JOINED TO A DNA PHOSPHORUS  
 CC AT ONE END OF THE ENZYME-SEVERED DNA STRAND.  
 CC -1- SIMILARITY: BELONGS TO PROKARYOTIC TYPE I/III TOPOISOMERASE  
 CC FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC DR EMBL; AEO04726; AAC06399.1; -  
 CC DR HSPSP; P06612; IECL.  
 CC DR InterPro; IPR002936; DNAPrim\_toprim.  
 CC DR InterPro; IPR003601; DNATopi\_ATP\_bind.  
 CC DR InterPro; IPR003602; DNATopi\_DNA\_bind.  
 CC DR InterPro; IPR00380; Prok\_Episomase.  
 CC DR Pfam; PF01131; Topoisom\_bac; 1.  
 CC DR Pfam; PF01396; zf-C4\_Topoiso; 2.  
 CC DR Pfam; PF01751; Toprim; 1.  
 CC DR PRINTS; PR00417; PRTPISMRASE1.  
 CC DR SMART; SM00437; TOP1AC; 1.  
 CC DR SMART; SM00436; TOP1BC; 1.  
 CC DR SMART; SM00493; TOPRIM; 1.  
 CC DR TIGRFAMs; TIGR01051; topA\_bact; 1.  
 CC DR PROSITE; PS00396; TOPOISOMERASE\_I\_PROK; 1.  
 CC DR Isomerase; topoisomerase; DNA-binding; Zinc-finger; Metal-binding;  
 CC Repeat; Complete proteome.  
 CC FT ZN\_FING 602 633 C4-TYPE 1.  
 CC FT ZN\_FING 664 691 C4-TYPE 2.  
 CC FT ZN\_FING 713 738 C4-TYPE 3.  
 CC FT ACT\_SITE 324 324 DNA\_CLEAVAGE (BY SIMILARITY).  
 CC SQ SEQUENCE 868 AA; 97282 MW; 5EB4B483366C98B CRC64;  
 CC  
 CC Query Match 83.3%; Score 30; DB 1; Length 868;  
 CC Best Local Similarity 100.0%; Pred. No. 59;  
 CC Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 CC  
 CC QY 2 ETRAPL 7  
 CC Db 793 ETRAPL 798  
 CC  
 CC RESULT 6  
 CC ID TOP1\_VIBCH STANDARD; PRT; 876 AA.  
 CC AC Q9KRB2;  
 CC DT 16-OCT-2001 (Rel. 40, Created)  
 CC DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 CC DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 CC DE DNA topoisomerase I (EC 5.99.1.2) (Omega-protein) (Relaxing enzyme)  
 CC DE (Unwisting enzyme) (Swivelase).  
 CC DE TOPA OR VC1730.  
 CC OS Vibrio cholerae.  
 CC OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.  
 CC NCBI\_TaxID=666;  
 CC RN NCI [1]  
 CC RP SEQUENCE FROM N.A.  
 CC RX STRAIN=El Tor N16961 / Serotype O1;  
 CC MEDLINE=20406833; PubMed=10952301;  
 CC RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,  
 CC Dodson R.J., Halt D.H., Hickey E.K., Peterson J.D., Umayam L.A.,  
 CC Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,  
 CC Ermlolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,  
 CC McDonald L., Uitterback T., Fleischmann R.D., Nierman W.C., White O.,  
 CC Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,  
 CC Fraser C.M.;  
 CC "DNA sequence of both chromosomes of the cholera pathogen Vibrio  
 CC cholerae.";  
 CC Nature 406:477-483(2000).  
 CC RL Nature 406:477-483(2000).  
 CC CC -1- FUNCTION: THE REACTION CATALYZED BY TOPOISOMERASES LEADS TO THE

CC CONVERSION OF ONE TOPOLOGICAL ISOMER OF DNA TO ANOTHER.  
 CC -1- CATALYTIC ACTIVITY: ATP-independent breakage of single-stranded  
 CC DNA, followed by passage and rejoining.  
 CC -1- SUBUNIT: MONOMER (BY SIMILARITY).  
 CC -1- MISCELLANEOUS: WHEN A TOPOISOMERASE TRANSIENTLY BREAKS A DNA  
 CC BACKBONE BOND, IT SIMULTANEOUSLY FORMS A PROTEIN-DNA LINK, IN  
 CC WHICH A TROSYL OXYGEN IN THE ENZYME IS JOINED TO A DNA PHOSPHORUS  
 CC AT ONE END OF THE ENZYME-SEVERED DNA STRAND.  
 CC -1- SIMILARITY: BELONGS TO PROKARYOTIC TYPE I/III TOPOISOMERASE  
 CC FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AEO04251; AAF94880.1; -.  
 CC HSSP; P06612; 1ECL.  
 CC TIGR; VC1730; -.  
 DR InterPro; IPR002936; DNAPrim\_toprim.  
 DR InterPro; IPR003601; DNACop1\_ATP\_bind.  
 DR InterPro; IPR003602; DNACop1\_DNA\_bind.  
 DR InterPro; IPR000380; Prok\_topoisomerase.  
 DR Pfam; PF01131; Topoisom\_bac; 1.  
 DR Pfam; PF01396; zf-C4\_Topoisom; 2.  
 DR Pfam; PF01751; Toprim; 1.  
 DR PRINTS; PR00417; PRTPISMASE1.  
 DR SMART; SM00437; TOPIAC; 1.  
 DR SMART; SM00436; TOPIBC; 1.  
 DR SMART; SM00493; TOPRIM; 1.  
 DR TIGR; TIGR01051; topa\_bact; 1.  
 DR PROSITE; PS00386; TOPOISOMERASE\_I\_PROK; 1.  
 DR Isomerase; Topoisomerase; DNA-binding; Zinc-finger; Metal-binding;  
 KW Repeat; Complete proteome.  
 FT ZN\_FING 668 695 C4-TYPE 1.  
 FT ZN\_FING 717 742 C4-TYPE 2.  
 FT ACT\_SITE 327 337 DNA\_CLEAVAGE (BY SIMILARITY).  
 FT SEQUENCE 876 AA; 97909 MW; 8A5BC9148CA1B753 CRC64;  
 SO  
 Query Match 83.3%; Score 30; DB 1; Length 876;  
 Best Local Similarity 100.0%; Pred. No. 60;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RL Virology 194:173-184(1993).  
 CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE  
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.  
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.  
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral  
 CC precursor polyprotein, commonly with Asp or Glu in the p6  
 CC position, Cys or Thr in p1 and Ser or Ala in p1'.  
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
 CC {RNA} (N).  
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 CC LIPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
 CC PROTEIN C AND RNA.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; L06436; AAA02739.1; -.  
 DR PIR; A46105; A46105.  
 DR HSSP; P14336; 1SVB.  
 DR MEROPS; S07.UPM; -.  
 DR InterPro; IPR001410; DEAD.  
 DR InterPro; IPR000069; Flavi\_M.  
 DR InterPro; IPR001157; Flavi\_NS1.  
 DR InterPro; IPR000752; Flavi\_NS2A.  
 DR InterPro; IPR000487; Flavi\_NS2B.  
 DR InterPro; IPR000404; Flavi\_NS4A.  
 DR InterPro; IPR001528; Flavi\_NS4B.  
 DR InterPro; IPR000208; Flavi\_NS5.  
 DR InterPro; IPR001122; Flavi\_capsidC.  
 DR InterPro; IPR000336; Flavi\_glycoprote.  
 DR InterPro; IPR001850; Flavi\_helicase.  
 DR InterPro; IPR002535; Flavi\_propep.  
 DR InterPro; IPR002877; Flavi.  
 DR InterPro; IPR001650; Helicase\_C.  
 DR Pfam; PF002711; helicase\_C; 1.  
 DR Pfam; PF00869; Flavi\_glycoprot; 1.  
 DR Pfam; PF00948; Flavi\_NS1; 1.  
 DR Pfam; PF00949; Flavi\_helicase; 1.  
 DR Pfam; PF00972; Flavi\_NS5; 1.  
 DR Pfam; PF01002; Flavi\_NS2B; 1.  
 DR Pfam; PF01003; Flavi\_capsid; 1.  
 DR Pfam; PF01004; Flavi\_M; 1.  
 DR Pfam; PF01005; Flavi\_NS2A; 1.  
 DR Pfam; PF01349; Flavi\_NS4B; 1.  
 DR Pfam; PF01350; Flavi\_NS4A; 1.  
 DR Pfam; PF01570; Flavi\_propep; 1.  
 DR Pfam; PF01728; Flavi; 1.  
 DR Pfam; PF02832; Flavi\_glycop\_C; 1.  
 DR PRODOM; PD001496; Flavi\_NS1; 1.  
 DR PRODOM; PD001556; Flavi\_glycoprote; 1.  
 DR SMART; SM00490; HELIC; 1.  
 KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;  
 KW Core protein; Coat protein; Envelope protein; Hydrolase; Helicase;  
 KW ATP-binding; Transmembrane; Nonstructural protein.  
 FT CHAIN 3 114  
 FT PROPEP 115 203  
 FT CHAIN 204 278  
 FT CHAIN 279 775  
 FT CHAIN 776 1128  
 FT CHAIN 1129 1358  
 FT CHAIN 1359 1489  
 FT CHAIN 1490 2111  
 FT CHAIN 2112 2250  
 FT CHAIN 2261 2512  
 FT CHAIN 2513 3415  
 FT TRANSMEM 101 114  
 FT TRANSMEM 244 262  
 ENVELOPE GLYCOPROTEIN M.  
 MAJOR ENVELOPE PROTEIN E.  
 NONSTRUCTURAL PROTEIN NS1.  
 NONSTRUCTURAL PROTEIN NS2A.  
 NONSTRUCTURAL PROTEIN NS2B.  
 PROTEASE/HELICASE (NS3).  
 NONSTRUCTURAL PROTEIN NS4A.  
 NONSTRUCTURAL PROTEIN NS4B.  
 RNA-DIRECTED RNA POLYMERASE (NS5).  
 HYDROPHOBIC SIGNAL SEQUENCE (POTENTIAL).  
 POTENTIAL.

```

FT TRANSMEM 264 278 HYDROPHOBIC SIGNAL SEQUENCE (POTENTIAL).
FT TRANSMEM 727 748 POTENTIAL.
FT TRANSMEM 754 775 POTENTIAL.
FT TRANSMEM 1137 1153 POTENTIAL.
FT TRANSMEM 1160 1179 POTENTIAL.
FT TRANSMEM 1294 1313 POTENTIAL.
FT TRANSMEM 1385 1403 POTENTIAL.
FT TRANSMEM 1453 1473 POTENTIAL.
FT TRANSMEM 2161 2184 POTENTIAL.
FT TRANSMEM 2191 2211 POTENTIAL.
FT TRANSMEM 2244 2260 HYDROPHOBIC SIGNAL SEQUENCE (POTENTIAL).
FT TRANSMEM 2347 2367 POTENTIAL.
FT TRANSMEM 2434 2454 POTENTIAL.
FT NP_BIND 1688 1695 ATP (BY SIMILARITY).
FT SITE 1780 1783 DEAH BOX.
FT DISULFID 281 308 BY SIMILARITY.
FT DISULFID 338 394 BY SIMILARITY.
FT DISULFID 352 383 BY SIMILARITY.
FT DISULFID 370 399 BY SIMILARITY.
FT DISULFID 464 568 BY SIMILARITY.
FT DISULFID 585 617 BY SIMILARITY.
FT CARBOHYD 142 142 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 432 432 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 860 860 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 983 983 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 999 999 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 3415 AA; 378564 MW; E71092FE6409F46 CRC64;

```

```

Query Match 83.3%; Score 30; DB 1; Length 3415;
Best Local Similarity 71.4%; Pred. No. 2.7e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```

```

Oy 1 GETRAFL 7
Db 2758 GETRGPI 2764

```

## RESULT 8

```

YTXM_BACSU STANDARD; PRT; 274 AA.
ID YTXM_BACSU STANDARD; PRT; 274 AA.
AC P23974; O34312;
DC 01-MAR-1992 (Rel. 21, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Putative esterase ytxm (EC 3.1.-.-).
GN YTXM.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_Taxid=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168 / RBL;
RA MEDLINE=96144257; PubMed=8566759;
RA Rowland B., Hill K., Miller P., Driscoll J.R., Taber H.W.;
RT "Structural organization of a Bacillus subtilis operon encoding
RT menaquinone biosynthetic enzymes.";
RL Gene 167:105-109(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA MEDLINE=98048467; PubMed=9387221;
RA Lapidus A., Galleron N., Sorokin A., Ehrlich S.D.;
RT "Sequencing and functional annotation of the Bacillus subtilis genes
RT in the 200 kb rmb-dnaB region.";
RL Microbiology 143:3431-3441(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azavedo V., Bertorello M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borries R., Boursier L., Brane A., Braun M., Biggelli S.C., Bron S.,
RA Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,

```

```

RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari B., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Funa S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaeser P., Goffeau A., Golligly E.J., Grandi G.,
RA Giuseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kaashara Y., Klaere-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Meinel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moesti D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Portwollik S., Prescott A.M.,
RA Prescan E., Puig P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadie Y.,
RA Sato T., Scallan E., Schleich S., Schroeter R., Scottone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Yamakoshi A., Tanaka T., Terpestra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenhof M., Vannier F., Vassarotti A.,
RA Viari A., Wambut R., Wedler E., Wedler H., Weisengger T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yamamoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis.";
RL Nature 390:249-256(1997).
CC -1 SIMILARITY: BELONGS TO THE LIP3/BCHO FAMILY OF LIPASES/ESTERASES.
CC -1 CAUTION: Ref.1 sequence differs from that shown due to many
CC frameshifts.
CC
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC
CC EMBL; M74538; AAC37015.1; ALT_FRAME.
CC EMBL; M74521; AAAS0400.1; ALT_FRAME.
CC EMBL; AF008220; AAC00225.1; -.
CC EMBL; Z99119; CAB15059.1; -.
CC PIR; S27511; S27511.
CC MEROPS; S33 UNM; -.
CC Subtilisin; B616685; ytxm.
CC InterPro; IPR003089; AB_hydrolase.
CC InterPro; IPR000073; Abhydrolase.
CC InterPro; IPR002410; Pro_aminopeptidase.
CC InterPro; IPR000379; Ser_estr_site.
CC Pfam; PF00561; abhydrolase_1.
CC Hypothetical protein; Hydrolase; Serine esterase; Complete proteome.
CC ACT_SITE 31 31 POTENTIAL.
CC ACT_SITE 99 99 CHARGE RELAY SYSTEM (BY SIMILARITY).
SQ SEQUENCE 274 AA; 30703 MW; 7C9C839D493D54CE CRC64;

```

```

Query Match 80.6%; Score 29; DB 1; Length 274;
Best Local Similarity 85.7%; Pred. No. 28;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

Oy 1 GETRAFL 7
Db 61 GETRAFL 67

```

## RESULT 9

```

ID GLYA_HELPY STANDARD; PRT; 416 AA.
AC P56089;
DC 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Serine hydroxymethyltransferase (EC 2.1.2.1) (Serine methylase)

```

```

DE (SHMT).
GN GLYA OR HP0183.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OK NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=26695 / ATCC 700392;
RX MEDLINE=97394467; PubMed=9252185;
RA Tomb J.-F., White O., Kervilange A.R., Clayton R.A., Sutton G.G.,
RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA Nelson K., Quackenbush U., Zhou L., Kirness E.F., Peterson S.,
RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA Berg D.B., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Matthey L., Wallin E.,
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Frazer C.M.,
RA Venter J.C.;
RT "The complete genome sequence of the gastric pathogen Helicobacter
RT pylori."
RL Nature 388:539-547(1997).
CC -1- FUNCTION: INTERCONVERSION OF SERINE AND GLYCINE.
CC -1- CATALYTIC ACTIVITY: 5,10-methylenetetrahydrofolate + glycine +
CC H(2)O = tetrahydrofolate + L-serine.
CC -1- COFACTOR: PYRIDOXAL PHOSPHATE (BY SIMILARITY).
CC -1- PATHWAY: KEY ENZYME IN THE BIOSYNTHESIS OF PURINES, LIPIDS,
CC HORMONES AND OTHER COMPONENTS.
CC -1- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE SHMT FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: AB000536; ADD07252.1; -.
DR HSSP: P00477; IDPO.
DR TIGR: HP0183; -.
DR InterPro: IPR001085; Gly_HyMettransf.
DR Pfam: PF00464; SHMT; 1.
DR PROSITE: PS00096; SHMT; 1.
DR Transfaser: Pyridoxal phosphate; One-carbon metabolism;
DR Complete proteome.
KW BINDING 226 226 PYRIDOXAL PHOSPHATE (BY SIMILARITY).
SQ SEQUENCE 416 AA; 45710 MW; 42BB0CD4433CD708 CRC64;
Query Match 80.6%; Score 29; DB 1; Length 416;
Best Local Similarity 83.3%; Pred. No. 45;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 GETRAP 6
DB 346 GETRAP 351

```

```

OK NCBI_TaxID=56636;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KL;
RX MEDLINE=99310339; PubMed=10382966;
RA Kawarabayashi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Anai A., Kosugi H.,
RA Hasegawa A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudo Y.,
RA Yamazaki J., Kishida N., Oguchi A., Aoki K.-I., Kubota K.,
RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
RT "Complete genome sequence of an aerobic hyper-thermophilic
RT crenarchaeon, Aeropyrum pernix K1."
RL DNA Res. 6:83-101(1999).
CC -1- FUNCTION: CONVERTS HMG-COA TO MEVALONATE (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: (R)-mevalonate + CoA + 2 NADP(+) = (S)-3-
CC hydroxy-3-methylglutaryl-CoA + 2 NADPH.
CC -1- PATHWAY: isoprenoid biosynthesis.
CC -1- SIMILARITY: BELONGS TO THE HMG-COA REDUCTASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: AB000662; BAA80874.1; -.
DR InterPro: IPR002202; HMG-CoA_red.
DR InterPro: IPR004554; HMG-CoA_R_NADP.
DR Pfam: PF00368; HMG-CoA_red; 1.
DR PRINTS: PR00071; HMGCOARDTASE.
DR TIGR: TIGR00533; HMG-CoA_R_NADP; 1.
DR PROSITE: PS00066; HMG-CoA_REDUCTASE_1; 1.
DR PROSITE: PS00318; HMG-CoA_REDUCTASE_2; 1.
DR PROSITE: PS01192; HMG-CoA_REDUCTASE_3; FALSE_NEG.
DR PROSITE: PS00065; HMG-CoA_REDUCTASE_4; 1.
DR Oxidoreductase; Isoprene biosynthesis; NADP; Complete proteome.
KW ACT SITE 109 109 BY SIMILARITY.
FT ACT SITE 315 315
FT ACT SITE 410 410 GENERAL BASE (BY SIMILARITY).
SQ SEQUENCE 421 AA; 44625 MW; 9F67AF57FA651822 CRC64;
Query Match 80.6%; Score 29; DB 1; Length 421;
Best Local Similarity 85.7%; Pred. No. 45;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 GETRAP 7
DB 137 GETRAP 143

```

```

RESULT 10
HMDH AERPE STANDARD; PRT; 421 AA.
AC 09YAS4;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE 3-hydroxy-3-methylglutaryl-coenzyme A reductase (EC 1.1.1.34) (HMG-CoA
DE reductase).
GN HMG A OR APT1869.
OS Aeropyrum pernix.
OC Archaea; Crenarchaeota; Thermoprotei; Desulfurococcaceae;
OC Desulfurococcaceae; Aeropyrum.

```

```

RESULT 11
NCAP BABVA STANDARD; PRT; 450 AA.
AC P15137;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-APR-1990 (Rel. 14, Last annotation update)
DE Nucleocapsid protein (Nucleoprotein).
DE N.
OS Rabies virus (strain AVO1).
OC Viruses; ssRNA negative-strand viruses; Mononegavirales;
OC Rhabdoviridae; Lyssavirus.
OX NCBI_TaxID=11293;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89150295; PubMed=3147698;
RA Roch O., Tordo N., Keth G.;
RT "Sequence of the 3386 3' nucleotides of the genome of the AVO1 strain
RT in transcription."

```

```
RL Biochimie 70:1019-1029(1988).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X13357; CA31733.1; -.
DR PIR; S07813; VHVNAV.
DR InterPro; IPR000448; Rhabd_nucleocap.
DR Pfam; PF00945; Rhabd_nucleocap; 1.
DR ProDom; PD002087; Rhabd_nucleocap; 1.
KM Nucleocapsid.
SQ SEQUENCE 450 AA; 50733 MW; 00493D2751305A76 CRC64;
Query Match 80.6%; Score 29; DB 1; Length 450;
Best Local Similarity 83.3%; Pred. No. 49;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 GETRAP 6
Db 397 GETRSP 402
RESULT 12
NCAP_RABVP STANDARD; PRT; 450 AA.
AC O08314;
DT 01-OCT-1994 (Rel. 30, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Nucleocapsid protein (Nucleoprotein).
GN N.
OS Rabies virus (strain Ontario fox).
OC Viruses; ssRNA negative-strand viruses; Mononegavirales;
OC Rhabdoviridae; Lyssavirus.
OX NCBI_TaxID=37132;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93260396; PubMed=8492088;
RA Nadin-Davis S.A.; Casey G.A.; Wandeler A.;
RT "Identification of regional variants of the rabies virus within the
RT Canadian province of Ontario."
RL J. Gen. Virol. 74:829-837(1993).
RN [2]
RP REVISIONS TO 444 AND 450.
RA Nadin-Davis S.A.;
RL Submitted (FEB-2001) to the EMBL/Genbank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L20673; AAA03482.2; -.
DR EMBL; L20675; AAA03484.2; -.
DR EMBL; L20676; AAA03485.2; -.
DR EMBL; L20672; AAA92762.1; -.
DR InterPro; IPR000448; Rhabd_nucleocap.
DR Pfam; PF00945; Rhabd_nucleocap; 1.
DR ProDom; PD002087; Rhabd_nucleocap; 1.
KM Nucleocapsid.
SQ SEQUENCE 450 AA; 50593 MW; D224ACADB7773B19 CRC64;
Query Match 80.6%; Score 29; DB 1; Length 450;
Best Local Similarity 83.3%; Pred. No. 49;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 GETRAP 6
Db 397 GETRSP 402
RESULT 13
NCAP_RABVP STANDARD; PRT; 450 AA.
AC P6025;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Nucleocapsid protein (Nucleoprotein).
GN N.
OS Rabies virus (strain Pasteur / PV).
OC Viruses; ssRNA negative-strand viruses; Mononegavirales;
OC Rhabdoviridae; Lyssavirus.
OX NCBI_TaxID=103929;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86176757; PubMed=3008096;
RA Tordo N., Poch O., Ermline A., Keith G.;
RT "Primary structure of leader RNA and nucleoprotein genes of the
RT rabies genome: segmented homology with VSV."
RL Nucleic Acids Res. 14:2671-2683(1986).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=86233343; PubMed=3459163;
RA Tordo N., Poch O., Ermline A., Keith G., Rougeon F.;
RT "Walking along the rabies genome: is the large G-L intergenic region
RT a remnant gene?";
RL Proc. Natl. Acad. Sci. U.S.A. 83:3914-3918(1986).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X03673; CAA27308.1; -.
DR EMBL; M13215; AAA47215.1; -.
DR EMBL; A14671; CAA01189.1; -.
DR EMBL; A14407; CAA01155.1; -.
DR PIR; D26275; VHVNAV.
DR InterPro; IPR000448; Rhabd_nucleocap.
DR Pfam; PF00945; Rhabd_nucleocap; 1.
DR ProDom; PD002087; Rhabd_nucleocap; 1.
KM Nucleocapsid.
SQ SEQUENCE 450 AA; 50605 MW; 670F9EC48985C3BB CRC64;
Query Match 80.6%; Score 29; DB 1; Length 450;
Best Local Similarity 83.3%; Pred. No. 49;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 GETRAP 6
Db 397 GETRSP 402
RESULT 14
NCAP_RABVS STANDARD; PRT; 450 AA.
AC P16285;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 01-AUG-1990 (Rel. 15, Last annotation update)
DE Nucleocapsid protein (Nucleoprotein).
GN N.
OS Rabies virus (strain SAD B19).
OC Viruses; ssRNA negative-strand viruses; Mononegavirales;
```

CC Rhabdoviridae; Lyssavirus.  
 OX NCBI\_TaxID=11300;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=90223994; PubMed=2139267;  
 RA Conzelmann K.-K., Cox J.H., Schneider L.G., Thiel H.-J.;  
 RT "Molecular cloning and complete nucleotide sequence of the attenuated  
 RL rabies virus SAD B19.";  
 CC Virology 175:485-499(1990).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; M31046; AAA47199.1; -.  
 DR PIR; A34746; VHVNSB.  
 DR InterPro; IPR000448; Rhabd\_nucleocap.  
 DR Pfam; PF00945; Rhabd\_nucleocap; 1.  
 DR ProDom; PD002087; Rhabd\_nucleocap; 1.  
 KW Nucleocapsid.  
 SQ SEQUENCE 450 AA; 50603 MW; 783BF3E01E7BE325 CRC64;

Query Match 80.6%; Score 29; DB 1; Length 450;  
 Best Local Similarity 83.3%; Pred. No. 49;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GETRAP 6  
 DB 397 GETRSP 402

RESULT 15  
 NCAP\_RABVU STANDARD; PRT; 450 AA.  
 ID NCAP\_RABVU 009110;  
 AC 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 01-FEB-1995 (Rel. 31, Last annotation update)  
 DE Nucleocapsid protein (Nucleoprotein).  
 GN N.  
 OS Rabies virus (strain Ontario skunk).  
 OC Viruses; ssRNA negative-strand viruses; Mononegavirales;  
 OC Rhabdoviridae; Lyssavirus.  
 OX NCBI\_TaxID=39005;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93260396; PubMed=8492088;  
 RA Nadin-Davis S.A., Casey G.A., Wadeler A.;  
 RT "Identification of regional variants of the rabies virus within the  
 RL Canadian province of Ontario.";  
 CC J. Gen. Virol. 74:829-837(1993).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; L20671; AAA47220.1; -.  
 DR InterPro; IPR000448; Rhabd\_nucleocap.  
 DR Pfam; PF00945; Rhabd\_nucleocap; 1.  
 DR ProDom; PD002087; Rhabd\_nucleocap; 1.  
 KW Nucleocapsid.  
 SQ SEQUENCE 450 AA; 50611 MW; D0F4CA9A7773C69 CRC64;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GETRAP 6  
 DB 397 GETRSP 402

Search completed: January 29, 2003, 14:00:24  
 Job time : 5.84615 secs

Query Match 80.6%; Score 29; DB 1; Length 450;  
 Best Local Similarity 83.3%; Pred. No. 49;

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2003, 13:56:21 ; Search time 19.5641 Seconds  
(without alignments)  
73.723 Million cell updates/sec

Title: US-09-807-949A-107  
Perfect score: 36  
Sequence: 1 GETRAFL 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

SPTRMBL\_21:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriaph:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	36	100.0	331	16	Q8UK92
2	36	100.0	1013	12	Q91LX9
3	33	91.7	403	10	Q8R245
4	33	91.7	949	5	Q9Y0H4
5	32	88.9	158	16	Q8Y101
6	32	88.9	220	12	Q69401
7	32	88.9	316	10	Q9C926
8	32	88.9	412	4	Q14184
9	32	88.9	412	11	P70169
10	32	88.9	413	10	Q8S282
11	32	88.9	449	11	Q9SM06
12	32	88.9	245	16	Q8XSM3
13	31	86.1	273	16	Q53671
14	31	86.1	345	5	Q9VXR2
15	31	86.1	525	2	Q9KXK8
16	31	86.1			Q9KXK8

17	31	86.1	526	2	Q93NK9	Q93nk9 yersinia en
18	31	86.1	1197	10	Q9SD06	Q9sd06 oryza sativ
19	31	86.1	1223	11	Q9QW33	Q9qw33 ratus sp.
20	30	83.3	106	2	Q51656	Q51656 paracoccus
21	30	83.3	121	16	Q9RJR8	Q9rjr8 streptomyce
22	30	83.3	128	16	Q8UDV1	Q8udv1 agrobacteri
23	30	83.3	168	16	Q9S2V7	Q9s2v7 streptomyce
24	30	83.3	188	2	Q9Z3P3	Q9z3p3 rhizobium s
25	30	83.3	227	8	Q9SB01	Q9sb01 glaphyrophe
26	30	83.3	227	8	Q9SAZ5	Q9saz5 pneumatocpe
27	30	83.3	246	16	Q9K3K2	Q9k3k2 streptomyce
28	30	83.3	270	17	Q972Y6	Q972y6 sulfolobus
29	30	83.3	274	16	Q8XXH8	Q8xxh8 ralsitonia s
30	30	83.3	278	16	Q93IU7	Q93iu7 streptomyce
31	30	83.3	344	16	Q8YCU3	Q8ycu3 bruceella me
32	30	83.3	352	11	Q9ESB4	Q9esb4 ratus norv
33	30	83.3	369	11	Q91ZP9	Q91zp9 mus musculu
34	30	83.3	405	16	Q9PCN8	Q9pcn8 xylella fas
35	30	83.3	441	16	Q8UBE6	Q8ube6 agrobacteri
36	30	83.3	479	10	Q9X1S2	Q9x1s2 arabidopsis
37	30	83.3	480	10	Q9X1S1	Q9x1s1 arabidopsis
38	30	83.3	481	10	Q9X1S3	Q9x1s3 arabidopsis
39	30	83.3	482	5	Q9U5Z3	Q9u5z3 caenorhabdi
40	30	83.3	482	10	Q94AL1	Q94al1 arabidopsis
41	30	83.3	487	10	Q9X1S4	Q9x1s4 arabidopsis
42	30	83.3	487	10	Q8VYL8	Q8vyl8 arabidopsis
43	30	83.3	492	16	Q93JK3	Q93jk3 streptomyce
44	30	83.3	529	10	Q9LJTS	Q9ljts arabidopsis
45	30	83.3	548	5	Q9Y049	Q9y049 caenorhabdi

#### ALIGNMENTS

RESULT 1	ID	Q8UK92	PRELIMINARY;	PRT;	331 AA.
AC	Q8UK92;				
DT	01-JUN-2002 (TREMBLrel. 21, Created)				
DT	01-JUN-2002 (TREMBLrel. 21, Last sequence update)				
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)				
DE	Acetyltransferase.				
GN	ATU5231 OR AGR PAT 325.				
OS	Agrobacterium tumefaciens (strain C58 / ATCC 33970).				
OG	Plasmid AT.				
OC	Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;				
OC	Rhizobiaceae; Rhizobium.				
OX	NCBI_TaxID=176299;				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=21608550; PubMed=11743193;				
RA	Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,				
RA	Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F., Jr., Woo L.,				
RA	Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sr.,				
RA	Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,				
RA	Kutyavin T., Levy R., Li M.-J., McClelland E., Palmeri A.,				
RA	Raymond C., Rouse G., Saenphimachak C., Wu Z., Romero P., Gordon D.,				
RA	Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Petry M.,				
RA	Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,				
RA	Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,				
RT	Nester E.W.;				
RT	"The genome of the natural genetic engineer Agrobacterium tumefaciens				
RT	C58.",				
KL	Science 294:2317-2323(2001).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=21608551; PubMed=11743194;				
RA	Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,				
RA	Currolo B., Goldman B.S., Cao Y., Askew M., Halling C., Mullin L.,				
RA	Houmel K., Gordon J., Vaudin M., Iatchouk O., Epp A., Liu F.,				
RA	Wolam C., Allinger M., Doughy D., Scott C., Lappas C., Marxelz B.,				
RA	Plangan C., Crowell C., Gursen J., Lomo C., Seer C., Strub G.,				
RA	Cielo C., Slater S.;				

```

RT "Genome sequence of the plant pathogen and biotechnology agent
RL Agrobacterium tumefaciens C58."
DR EMBL; AE008945; AALAS928.1; AUT_INIT.
DR EMBL; AE007893; AAK90602.1; -.
KW Plasmid; Transferase; Complete proteome.
SQ SEQUENCE 331 AA; 36623 MW; APEFD9D9479DD539 CRC64;

Query Match
Query Local Similarity 100.0%; Score 36; DB 16; Length 331;
Best Local Similarity 100.0%; Pred. No. 6.4;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 GETRAPL 7
Db 315 GETRAPL 321

RESULT 2
O9ILX9 PRELIMINARY; PRT; 1013 AA.
ID O9ILX9
AC O9ILX9;
DT 01-OCT-2000 (TRENBLrel. 15, Created)
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
DT 01-MAR-2002 (TRENBLrel. 20, Last annotation update)
DE DNA polymerase.
OS retroperitonal fibromatosis-associated herpesvirus.
OC viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Rhadinovirus.
ON NCBI_TaxID=111469;
RX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MNE442N;
RX MEDLINE=20240083; PubMed=10775636;
RA Schulz E.R., Rankin G.W. Jr., Blanc M.P., Raden B.W., Tsai C.C.,
RA Rose T.M.;
RT "Characterization of two divergent lineages of macaque rhadinoviruses
RT related to Kaposi's sarcoma-associated herpesvirus."
RL J. Virol. 74:4919-4928(2000).
CC -1- CATALYTIC ACTIVITY: N DEOXYNUCLEOSIDE TRIPHOSPHATE = N DIPHOSPHATE
CC + (DNA)(N).
CC -1- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-B FAMILY.
DR EMBL; AF204166; AAF81662.1; -.
DR InterPro; IPR002064; DNA_pol_B.
DR Pfam; PF00136; DNA_pol_B; 1.
DR Pfam; PF03104; DNA_pol_B_exo; 1.
DR PRINTS; PR00106; DNA_POLB.
DR SMART; SM00486; POLBc; 1.
DR PROSITE; PS00116; DNA_POLYMERASE_B; 1.
KW DNA replication; DNA-binding; DNA-directed DNA polymerase.
SQ SEQUENCE 1013 AA; 113915 MW; 39C1F845764A5B0 CRC64;

Query Match
Query Local Similarity 100.0%; Score 36; DB 12; Length 1013;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 GETRAPL 7
Db 92 GETRAPL 98

RESULT 3
O8R245 PRELIMINARY; PRT; 403 AA.
ID O8R245
AC O8R245;
DT 01-JUN-2002 (TRENBLrel. 21, Created)
DT 01-JUN-2002 (TRENBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE Lipase-like protein.
GN OJ1159_D09.24.
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.

```

```

OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 1, BAC
RT clone:OJ1159_D09.24."
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP003792; BAB89206.1; -.
SQ SEQUENCE 403 AA; 45035 MW; 522C4B1CC437F44C CRC64;

Query Match
Query Local Similarity 91.7%; Score 33; DB 10; Length 403;
Best Local Similarity 85.7%; Pred. No. 38;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 1 GETRAPL 7
Db 245 GETRAPV 251

RESULT 4
O9Y0H4 PRELIMINARY; PRT; 949 AA.
ID O9Y0H4
AC O9Y0H4;
DT 01-NOV-1999 (TRENBLrel. 12, Created)
DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE Su(dx) protein.
ON SU(DX) OR CG4244.
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachyera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
ON NCBI_TaxID=7227;
RX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10711132;
RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.B., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Man K.H., Doyle C., Baxter B.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Chewley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fowler C., Gabrielian A.E., Garcia N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harts N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howard T.J., Wei M.-H., Idegan C.,
RA Jaitai M., Kalish F., Karpen G.H., Ke Z., Kemmison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lascko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Matel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Murry D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Paclob J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wansarman D.A., Weinstein G.M., Weisenbach U.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,

```

RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Corneli M., Evans D.A.P., Mann R., Fostier M., Flaaza M.,  
 RA Monthong M., Aravanis-Tsakonas S., Baron M.;  
 RT "The *Drosophila melanogaster* Suppressor of deltex gene, a regulator of  
 RT the Notch receptor signalling pathway, is an E3 class ubiquitin  
 RT ligase.";  
 RL Genetics 152:0-0(1999).  
 DR EMBL; AF003584; AAF512.1; -;  
 DR EMBL; AF152865; AAD38975.1; -;  
 DR HSSP; Q13526; IPIN.  
 DR FBase; FBgn0003557; Su(dx).  
 DR InterPro; IPR000008; C2  
 DR InterPro; IPR000569; HECT\_domain.  
 DR InterPro; IPR001202; WW\_Rsp5\_WWP.  
 DR Pfam; PF00168; C2; 1.  
 DR Pfam; PF00632; HECT; 1.  
 DR Pfam; PF00397; WW; 4.  
 DR SMART; SM00239; C2; 1.  
 DR SMART; SM00119; HECT; 1.  
 DR SMART; SM00456; WW; 3.  
 DR PROSITE; PS50004; C2\_DOMAIN\_2; 2.  
 DR PROSITE; PS50237; HECT; 2.  
 DR PROSITE; PS01159; WW\_DOMAIN\_1; 3.  
 DR PROSITE; PS50020; WW\_DOMAIN\_2; 4.  
 SQ SEQUENCE 949 AA; 107966 MW; 74B17A8B05ACE6B CRC64;

Query Match 91.7%; Score 33; DB 5; Length 949;  
 Best Local Similarity 85.7%; Pred. No. 96;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GETRAPL 7  
 Db 236 GETRSP 242

RESULT 5  
 ID 08Y101 PRELIMINARY; PRT; 158 AA.  
 AC 08Y101;  
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
 DE Putative phage-related protein.  
 GN RSC0892 OR RS04521.  
 OS Ralstonia solanacearum (pseudomonas solanacearum).  
 OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;  
 OC Ralstonia.  
 NC NCB1\_TaxID=305;  
 RX MEDLINE=21681879; PubMed=11823852;  
 RX Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,  
 RA Arlat M., Billault A., Broctier P., Camus J.C., Catolico L.,  
 RA Chandler M., Choisme N., Claudel-Renard C., Cunnac S., Demange N.,  
 RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T.,  
 RA Signier P., Thébault P., Whalen M., Wincker P., Levy M.,  
 RA Weisenbach J., Boucher C.A.;  
 RT "Genome sequence of the plant pathogen *Ralstonia solanacearum*.";  
 RL Nature 415:497-502(2002).  
 DR EMBL; AL646061; CADI4594.1; -;  
 KW Complete proteome.  
 SQ SEQUENCE 158 AA; 17621 MW; 8FA7BCA1A39440DE CRC64;

Query Match 88.9%; Score 32; DB 16; Length 158;  
 Best Local Similarity 100.0%; Pred. No. 23;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 GETRAP 6

Db 50 GETRAP 55

RESULT 6  
 ID 069401 PRELIMINARY; PRT; 220 AA.  
 AC 069401;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE UL3.5.  
 GN UL3.5.  
 OS Pseudorabies virus.  
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 OC Alphaherpesvirinae; Varicelloviruses.  
 NC NCB1\_TaxID=10345;  
 RX MEDLINE=93381797; PubMed=8396663;  
 RA Dean H.U., Cheung A.K.;  
 RT "A 3' coterminal gene cluster in pseudorabies virus contains herpes  
 RT simplex virus UL1, UL2, and UL3 gene homologs and a unique UL3.5 open  
 RT reading frame.";  
 RL J. Virol. 67:5955-5961(1993).  
 DR EMBL; L13855; AAA16424.1; -;  
 SQ SEQUENCE 220 AA; 23670 MW; BAA2983EC16D95C5 CRC64;

Query Match 88.9%; Score 32; DB 12; Length 220;  
 Best Local Similarity 100.0%; Pred. No. 33;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GETRAP 6  
 Db 41 GETRAP 46

RESULT 7  
 ID 09C926 PRELIMINARY; PRT; 316 AA.  
 AC 09C926;  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE Hypothetical 34.7 kDa protein [AT3908600/FL1014.7].  
 GN FL1014.7.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC Eurosid II; Brassicales; Brassicaceae; Arabidopsids.  
 NC NCB1\_TaxID=3702;  
 RX MEDLINE=21016720; PubMed=11130713;  
 RX Salanoubat M., Lemcke K., Rieger M., Ansoerge W., Unefeld M.,  
 RA Peltmann B., Valle G., Bloeker H., Perez-Alonso M., Obermaier B.,  
 RA Delannoy M., Boutry M., Griwall L.A., Mache R., Puigdomenech P.,  
 RA De Simone V., Choisme N., Artiguenave F., Robert C., Broctier P.,  
 RA Wincker P., Catolico L., Weisenbach J., Saurin W., Quetier F.,  
 RA Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Benes V.,  
 RA Wurmisch B., Drzonek H., Erfle H., Jordan N., Bangert S.,  
 RA Wiedemann R., Kranz H., Voss H., Holland R., Brandt P., Nyakatura G.,  
 RA Vezzi A., D'Angelo M., Pallavicini A., Toppo S., Simionati B.,  
 RA Reichelt J., Scharte C., Kauer G., Loehner T.-H., Nordstiek G.,  
 RA Navarro P., Collado C., Perez-Perez A., Bargues M., Terol J., Clement J.,  
 RA Cooke R., Landie M., Berger-Blatro C., Fumelle B., Masuy D.,  
 RA de Haan M., Maarse A.C., Alcaraz J.-P., Cottet A., Casasuberta E.,  
 RA Monfort A., Argilou A., Flores M., Liguori R., Vitale D.,  
 RA Mennhaup G., Haase D., Schoof H., Rudd S., Zaccaria P., Mewes H.-W.,  
 RA Mayer K.F.X., Kaul S., Town C.D., Koo H.L., Tallon L.J., Jenkins J.,

```

RA Rooney T., Rizzo M., Walte A., Utechtack T., Fujii C.Y., Shea T.P.,
RA Creasy T.H., Haas B., Maiti R., Wu D., Peterson J., Van Aken S.,
RA Pal G., Miltischer J., Sellers P., Gill J.E., Feldblyum T.V.,
RA Preuss D., Lin X., Nieman W.C., Salzberg S.L., White O., Venter J.C.,
RA Fraser C.M., Kaneko T., Nakamura Y., Sato S., Kato T., Asamizu E.,
RA Sasamoto S., Kimura T., Ideawa K., Kawashima K., Kishida Y.,
RA Kiyokawa S., Kohara M., Matsumoto M., Taketani A.,
RA Nakayama S., Kikazaki N., Shinozaki S., Takeda C., Wada T.,
RA Matsumoto A., Yamada M., Yasuda M., Tabata S.,
RT "Sequence and analysis of chromosome 3 of the plant Arabidopsis
RT thaliana."
RL Nature 408:820-822(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Kim C.J., Chen H., Cheuk R., Koesema E., Meyers M.C., Banh J.,
RA Bowser L., Garnier P., Dale J.M., Goldsmith A.D., Hayashizaki Y.,
RA Ishida J., Jiang P.X., Jones T., Kamiya A., Karlin-Neumann G.,
RA Kawai J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M.,
RA Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,
RA Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,
RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
RA Ecker J.R.;
RT "Arabidopsis ORF clones."
RT Submitted (SEP-2001) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Kim C.J., Chen H., Cheuk R., Koesema E., Meyers M.C., Banh J.,
RA Bowser L., Garnier P., Dale J.M., Gibson H.A., Goldsmith A.D.,
RA Hayashizaki Y., Ishida J., Jiang P.X., Jones T., Kamiya A.,
RA Karlin-Neumann G., Kawai J., Lam B., Lee J.M., Lin J., Liu S.X.,
RA Miranda M., Narusaka M., Nguyen M., Onodera C.S., Palm C.J.,
RA Pham P.K., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A.,
RA Tang C.C., Toriumi M., Yamada K., Yu G., Yu S., Shinozaki K.,
RA Davis R.W., Theologis A., Ecker J.R.;
RT "Arabidopsis cDNA clones."
RT Submitted (JUL-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL; AC012562; AAC51364.1; -.
DR EMBL; AY045687; AA106975.1; -.
DR EMBL; AY045678; AAK74036.1; -.
KW Hypothetical protein.
SQ SEQUENCE 316 AA; 34732 MW; 5B54FCF59A5B5B5 CRC64;

Query Match 88.9%; Score 32; DB 10; Length 316;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GETRAP 6
Db 294 GETRAP 299

RESULT 8
Q14184 PRELIMINARY; PRT; 412 AA.
AC Q14184;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE Doc2 beta.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96125347; PubMed=8554557;
RA Satauchi G., Orita S., Maeda M., Igatah H., Takai Y.;
RT "Molecular cloning of an isoform of Doc2 having two C2-like domains."
RL Biochem. Biophys. Res. Commun. 217:1053-1061(1996).
DR EMBL; D70830; BA01107.1; -.
DR HSSP; P21707; IRSY.
DR InterPro; IPR000008; C2.
DR InterPro; IPR002149; LRI.

Query Match 88.9%; Score 32; DB 10; Length 316;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GETRAP 6
Db 294 GETRAP 299

```

```

DR InterPro; IPR001565; Synaptotagmin.
DR Pfam; PF00168; C2; 2.
DR PRINTS; PR00360; C2DOMAIN.
DR PRINTS; PR00399; SYNAPTOTAGMN.
DR SMART; SM00239; C2; 2.
DR PROSITE; PS00499; C2_DOMAIN_1; 2.
DR PROSITE; PS50004; C2_DOMAIN_2; 2.
SQ SEQUENCE 412 AA; 45949 MW; 679592A874542ED3 CRC64;

Query Match 88.9%; Score 32; DB 4; Length 412;
Best Local Similarity 85.7%; Pred. No. 65;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GETRAP 7
Db 229 GETRAP 235

RESULT 9
Q14184 PRELIMINARY; PRT; 412 AA.
AC Q14184;
DT 01-FEB-1997 (TReMBLrel. 02, Created)
DT 01-FEB-1997 (TReMBLrel. 02, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE Doc2beta.
GN Doc2B.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97058315; PubMed=8902635;
RA Kojima T., Fukuda M., Aruga J., Mikoshiba K.;
RT "Calcium-dependent phospholipid binding to the C2A domain of a
RT ubiquitous form of double C2 protein (doc2beta).";
RL J. Biochem. 120:671-676(1996).
DR EMBL; D85037; BA012714.1; -.
DR HSSP; P21707; IRSY.
DR MGD; MGI:1100497; Doc2b.
DR InterPro; IPR000008; C2.
DR InterPro; IPR002149; LRI.
DR InterPro; IPR001565; Synaptotagmin.
DR Pfam; PF00168; C2; 2.
DR PRINTS; PR00360; C2DOMAIN.
DR PRINTS; PR00399; SYNAPTOTAGMN.
DR SMART; SM00239; C2; 2.
DR PROSITE; PS00499; C2_DOMAIN_1; 2.
DR PROSITE; PS50004; C2_DOMAIN_2; 2.
SQ SEQUENCE 412 AA; 45839 MW; BEC1480BDE91724 CRC64;

Query Match 88.9%; Score 32; DB 11; Length 412;
Best Local Similarity 85.7%; Pred. No. 65;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GETRAP 7
Db 229 GETRAP 235

RESULT 10
Q14184 PRELIMINARY; PRT; 412 AA.
AC Q14184;
DT 01-FEB-1997 (TReMBLrel. 02, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE Doc2B.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

```

OK NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=BRAIN;  
 RA Verlage M., Sudhof T.C.;  
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; U70778; AAA47747.2; -.  
 DR HSSP; P21707; IRSY.  
 DR InterPro; IPR000008; C2.  
 DR InterPro; IPR002149; LRI.  
 DR InterPro; IPR001565; Synaptotagmin.  
 DR Pfam; PF00168; C2; 2.  
 DR PRINTS; PR00360; C2DOMAIN.  
 DR PRINTS; PR00399; SYNAPTOTAGMIN.  
 DR SMART; SM00239; C2; 2.  
 DR PROSITE; PS00499; C2\_DOMAIN\_1; 2.  
 DR PROSITE; PS00004; C2\_DOMAIN\_2; 3.  
 SQ SEQUENCE 412 AA; 45841 MW; 61595BC6866F1012 CRC64;

Query Match 88.9%; Score 32; DB 11; Length 412;  
 Best Local Similarity 85.7%; Pred. No. 65;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GETRAPL 7  
 DB 229 GETRVPL 235

RESULT 11

Q8S282 PRELIMINARY; PRT; 413 AA.  
 AC Q8S282;  
 DT 01-JUN-2002 (TREMBLrel. 21, Created)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)  
 DE Putative MYB transcription factor.  
 GN P0414B03.21.  
 OS Oryza sativa (japonica cultivar-group).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzaceae; Oryza.  
 OC NCBI\_TaxID=39947;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. NIPPONBARI;  
 RA Sasaki T., Matsunoto T., Yamamoto K.;  
 RT "Oryza sativa nipponbare(GAS) genomic DNA, chromosome 1, PAC  
 RT clone:PD414E03."  
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AP003242; BAB89519.1; -.  
 SQ SEQUENCE 413 AA; 44329 MW; 173D1B6C50C245A CRC64;

Query Match 88.9%; Score 32; DB 10; Length 413;  
 Best Local Similarity 85.7%; Pred. No. 66;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GETRAPL 7  
 DB 61 GETRVPL 67

RESULT 12

O99M06 PRELIMINARY; PRT; 449 AA.  
 AC O99M06;  
 DT 01-JUN-2001 (TREMBLrel. 17, Created)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Hypothetical 50.7 kDa protein (fragment).  
 OS Mus musculus (mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OC NCBI\_TaxID=10090;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Strausberg R.;  
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC002138; AAH02138.1; -.  
 DR MEROPS; S09.054; -.  
 DR InterPro; IPR000379; Ser\_estrs\_site.  
 KM Hypothetical protein.  
 FT NON TER 1  
 SQ SEQUENCE 449 AA; 50723 MW; 973C812F92F6797A CRC64;

Query Match 88.9%; Score 32; DB 11; Length 449;  
 Best Local Similarity 100.0%; Pred. No. 72;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GETRAPL 6  
 DB 37 GETRVPL 42

RESULT 13

O8XSM3 PRELIMINARY; PRT; 245 AA.  
 AC O8XSM3;  
 DT 01-MAR-2002 (TREMBLrel. 20, Created)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
 DE Putative FIRIN-like protein.  
 GN RSP0444 OR RS00903.  
 OS Ralstonia solanacearum (Pseudomonas solanacearum).  
 OC Ralstonia solanacearum (Pseudomonas solanacearum).  
 OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;  
 OC Ralstonia.  
 OC NCBI\_TaxID=305;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=GM11000;  
 RX MEDLINE=21681879; PubMed=11823852;  
 RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,  
 RA Ariat M., Billault A., Brottier P., Camus J.C., Cactolico L.,  
 RA Chandler M., Cholise N., Claudel-Renard C., Cunac S., Demange N.,  
 RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T.,  
 RA Siquier P., Thebaud P., Whalen M., Wincker P., Levy M.,  
 RA Weisenbach J., Boucher C.A.;  
 RT "Genome sequence of the plant pathogen Ralstonia solanacearum";  
 RL Nature 415:497-502 (2002).  
 DR EMBL; AL640768; CAD17595.1; -.  
 KM Plasmid; Complete proteome.  
 SQ SEQUENCE 245 AA; 26062 MW; 436F2D1843FD60DC CRC64;

Query Match 86.1%; Score 31; DB 16; Length 245;  
 Best Local Similarity 85.7%; Pred. No. 63;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GETRAPL 7  
 DB 151 GETRVPL 157

RESULT 14

O53671 PRELIMINARY; PRT; 273 AA.  
 AC O53671;  
 DT 01-JUN-1998 (TREMBLrel. 06, Created)  
 DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
 DE Hypothetical protein Rv0249c.  
 GN RV0249C OR MTU034.15C OR MTU263.  
 OS Mycobacterium tuberculosis.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OC NCBI\_TaxID=1773;  
 RN [1]

```

RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gao S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrall B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / OSHKOSH;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., Deboy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson M.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Kouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains."
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL021929; CAAL7341.1; -
DR EMBL: AB006934; AAK44481.1; -
DR TIGR: MT0263; -
DR TubercuList; RV0249c; -
RW Hypothetical protein; Complete proteome.
SQ SEQUENCE 273 AA; 31077 MW; FA881203425B91EA CRC64;

Query March 86.1%; Score 31; DB 16; Length 273;
Best Local Similarity 85.7%; Pred. No. 70;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GETRAPL 7
Db 143 GETRAPL 149

RESULT 15
Q9VXR2 PRELIMINARY; PRT; 345 AA.
AC Q9VXR2;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-MAR-2002 (TRENBLrel. 20, Last annotation update)
DE CG8191 protein (REL1765Sp).
GN CG8191.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brendon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Ffiankoch C., Baldwin D.,
RA Ballew R.M., Baas A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Butris K.C., Busam D.A., Butler H., Cadieu E., Genter A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

```

```

RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hotin D., Houston K.A., Howland T.J., Wei M.-H., Ilegem C.,
RA Jaitai M., Kalush F., Kapen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laake P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Melnikov G., Mlishina N.V., Mobarry C., Morris J., Mostreffi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusseken D.R., Paclet J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler P., Smith T.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Stryckas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA George R., Gonzalez M., Guarin H., Krommiller B., Li P., Liao G.,
RA Miranda A., Mungall C.J., Nunco J., Paclet J., Paragas V., Park S.,
RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA Celisner S.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB003500; AAR48496.1; -
DR EMBL: AY071134; AAL48756.1; -
DR FlyBase; FBgn0030675; CG8191.
SQ SEQUENCE 345 AA; 39831 MW; 0613409F8F007B9C CRC64;

Query March 86.1%; Score 31; DB 5; Length 345;
Best Local Similarity 85.7%; Pred. No. 91;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GETRAPL 7
Db 109 GETRAPL 115

Search completed: January 29, 2003, 14:02:21
Job time : 21.5641 secs

```

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using SW model

Run on: January 29, 2003, 13:57:26 ; Search time 8.61539 Seconds  
(without alignments)  
23.906 Million cell updates/sec

Title: US-09-807-949a-107  
Perfect score: 36  
Sequence: 1 GETRAPL 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : Issued Patents AA:

1: /cgn2\_6/prodata/1/aa/5A\_COMB.pep:\*  
2: /cgn2\_6/prodata/1/aa/5B\_COMB.pep:\*  
3: /cgn2\_6/prodata/1/aa/6A\_COMB.pep:\*  
4: /cgn2\_6/prodata/1/aa/6B\_COMB.pep:\*  
5: /cgn2\_6/prodata/1/aa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/prodata/1/aa/backfillseq1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	36	100.0	7	4	US-09-178-115-107
2	36	100.0	7	4	US-09-177-776-107
3	36	100.0	7	4	US-09-084-605B-29
4	36	100.0	7	4	US-09-438-150-6
5	31	86.1	7	4	US-09-178-115-108
6	31	86.1	7	4	US-09-177-776-108
7	30	83.3	7	4	US-09-178-115-109
8	30	83.3	7	4	US-09-177-776-109
9	28	77.8	54	2	US-08-456-647B-51
10	28	77.8	54	2	US-08-237-401A-51
11	28	77.8	367	4	US-09-380-326-9
12	28	77.8	514	4	US-09-413-814-82
13	28	77.8	514	4	US-09-413-814-94
14	28	77.8	571	2	US-08-796-414B-1
15	28	77.8	771	3	US-09-121-964-9
16	28	77.8	933	2	US-08-330-1
17	28	77.8	933	5	PCT-US93-03837-1
18	28	77.8	1311	1	US-08-340-011-5
19	28	77.8	1311	3	US-08-901-710-5
20	28	77.8	1338	3	US-08-750-141A-3
21	27	75.0	71	4	US-08-963-851-16
22	27	75.0	247	4	US-09-370-831-112
23	27	75.0	279	4	US-09-134-001C-3260
24	27	75.0	295	4	US-09-134-001C-2850
25	27	75.0	359	1	US-08-181-271A-98
26	27	75.0	359	1	US-08-449-315-98
27	27	75.0	359	1	US-08-444-803-98

28	27	75.0	359	1	US-08-047-413-15	Sequence 15, Appl
29	27	75.0	359	1	US-08-449-043-98	Sequence 98, Appl
30	27	75.0	359	1	US-08-456-265A-98	Sequence 98, Appl
31	27	75.0	359	1	US-08-455-416-98	Sequence 98, Appl
32	27	75.0	359	1	US-08-455-244-98	Sequence 98, Appl
33	27	75.0	359	1	US-08-454-876-98	Sequence 98, Appl
34	27	75.0	359	2	US-08-457-364-98	Sequence 98, Appl
35	27	75.0	359	2	US-08-456-263-98	Sequence 98, Appl
36	27	75.0	359	2	US-08-456-240-98	Sequence 98, Appl
37	27	75.0	359	2	US-08-455-736-98	Sequence 98, Appl
38	27	75.0	359	2	US-08-971-217-98	Sequence 98, Appl
39	27	75.0	359	3	US-08-229-050-15	Sequence 15, Appl
40	27	75.0	359	3	US-08-801-563-15	Sequence 15, Appl
41	27	75.0	359	4	US-09-350-600-98	Sequence 98, Appl
42	27	75.0	383	4	US-09-206-800-6	Sequence 6, Appl
43	27	75.0	383	4	US-09-206-800-7	Sequence 7, Appl
44	27	75.0	383	4	US-09-206-800-8	Sequence 8, Appl
45	27	75.0	635	1	US-07-832-855-2	Sequence 2, Appl

#### ALIGNMENTS

```

RESULT 1
US-09-178-115-107
; Sequence 107, Application US/09178115
; Patent No. 6297041
; GENERAL INFORMATION:
; APPLICANT: Zavada, Jan
; APPLICANT: Pastorekova, Silvia
; APPLICANT: Pastorek, Jaromir
; TITLE OF INVENTION: MN Gene and Protein
; FILE REFERENCE: D-0021 5A
; CURRENT APPLICATION NUMBER: US/09178, 115
; EARLIER FILING DATE: 1998-10-23
; EARLIER APPLICATION NUMBER: 09/177, 776
; EARLIER FILING DATE: 1998-10-23
; EARLIER APPLICATION NUMBER: 08/787, 739
; EARLIER FILING DATE: 1997-01-24
; EARLIER APPLICATION NUMBER: 08/485, 049
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 08/486, 756
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 08/477, 504
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 08/481, 658
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 08/485, 862
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 08/485, 863
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 08/487, 077
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 08/260, 190
; EARLIER FILING DATE: 1994-06-15
; EARLIER APPLICATION NUMBER: 08/177, 093
; EARLIER FILING DATE: 1993-12-30
; EARLIER APPLICATION NUMBER: 07/964, 589
; EARLIER FILING DATE: 1992-10-21
; EARLIER APPLICATION NUMBER: PV-709-92
; EARLIER FILING DATE: 1992-03-11
; NUMBER OF SEQ ID NOS: 116
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 107
; LENGTH: 7
; TYPE: PRT
; ORGANISM: HUMAN
US-09-178-115-107
Query Match 100.0%; Score 36; DB 4; Length 7;
Best Local Similarity 100.0%; Pred. No. 1.9e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 1 GETRAPL 7  
| | | | |  
Db 1 GETRAPL 7

## RESULT 2

US-09-177-776-107  
; Sequence 107, Application US/09177776A  
; Patent No. 6297051  
; GENERAL INFORMATION:  
; APPLICANT: Zavada, Jan  
; APPLICANT: Pastorekova, Silvia  
; APPLICANT: Pastorek, Jaromir  
; TITLE OF INVENTION: MN Gene and Protein  
; FILE REFERENCE: D-0021.5A  
; CURRENT APPLICATION NUMBER: US/09/177,776A  
; EARLIER FILING DATE: 1998-10-23  
; EARLIER APPLICATION NUMBER: 08/787,739  
; EARLIER FILING DATE: 1997-01-24  
; EARLIER APPLICATION NUMBER: 08/485,049  
; EARLIER FILING DATE: 1995-06-07  
; EARLIER APPLICATION NUMBER: 08/486,756  
; EARLIER FILING DATE: 1995-06-07  
; EARLIER APPLICATION NUMBER: 08/477,504  
; EARLIER FILING DATE: 1995-06-07  
; EARLIER APPLICATION NUMBER: 08/481,658  
; EARLIER FILING DATE: 1995-06-07  
; EARLIER APPLICATION NUMBER: 08/485,862  
; EARLIER FILING DATE: 1995-06-07  
; EARLIER APPLICATION NUMBER: 08/485,863  
; EARLIER FILING DATE: 1995-06-07  
; EARLIER APPLICATION NUMBER: 08/487,077  
; EARLIER FILING DATE: 1995-06-07  
; EARLIER APPLICATION NUMBER: 08/260,190  
; EARLIER FILING DATE: 1994-06-15  
; EARLIER APPLICATION NUMBER: 08/177,093  
; EARLIER FILING DATE: 1993-12-30  
; EARLIER APPLICATION NUMBER: 07/964,589  
; EARLIER FILING DATE: 1992-10-21  
; EARLIER APPLICATION NUMBER: PV-709-92  
; EARLIER FILING DATE: 1992-03-11  
; NUMBER OF SEQ ID NOS: 116  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 107  
; LENGTH: 7  
; TYPE: PRT  
; ORGANISM: HUMAN  
US-09-177-776-107

Query Match 100.0%; Score 36; DB 4; Length 7;  
Best Local Similarity 100.0%; Pred. No. 1.9e+05;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GETRAPL 7  
| | | | |  
Db 1 GETRAPL 7

## RESULT 3

US-09-084-605B-29  
; Sequence 29, Application US/09084605B  
; Patent No. 6329501  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Bruce F.  
; APPLICANT: Samoilova, Tatiana  
; TITLE OF INVENTION: Methods and Compositions for Targeting  
; FILE REFERENCE: 5721-8  
; CURRENT APPLICATION NUMBER: US/09/084,605B  
; CURRENT FILING DATE: 1998-05-26  
; NUMBER OF SEQ ID NOS: 29  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 29

; LENGTH: 7  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Phage display library peptides  
US-09-084-605B-29

Query Match 100.0%; Score 36; DB 4; Length 7;  
Best Local Similarity 100.0%; Pred. No. 1.9e+05;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GETRAPL 7  
| | | | |  
Db 1 GETRAPL 7

## RESULT 4

US-09-438-150-6  
; Sequence 6, Application US/09438150  
; Patent No. 6399575  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Bruce F.  
; APPLICANT: Samoilova, Tatiana I.  
; APPLICANT: Baker, Henry J.  
; TITLE OF INVENTION: Methods and Compositions for Targeting  
; FILE REFERENCE: 5721-13  
; CURRENT APPLICATION NUMBER: US/09/438,150  
; CURRENT FILING DATE: 1999-11-10  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 6  
; LENGTH: 7  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Phage display library peptides  
US-09-438-150-6

Query Match 100.0%; Score 36; DB 4; Length 7;  
Best Local Similarity 100.0%; Pred. No. 1.9e+05;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GETRAPL 7  
| | | | |  
Db 1 GETRAPL 7

## RESULT 5

US-09-178-115-108  
; Sequence 108, Application US/09178115  
; Patent No. 6297041  
; GENERAL INFORMATION:  
; APPLICANT: Zavada, Jan  
; APPLICANT: Pastorekova, Silvia  
; APPLICANT: Pastorek, Jaromir  
; TITLE OF INVENTION: MN Gene and Protein  
; FILE REFERENCE: D-0021.5A  
; CURRENT APPLICATION NUMBER: US/09/178,115  
; CURRENT FILING DATE: 1998-10-23  
; EARLIER APPLICATION NUMBER: 09/177,776  
; EARLIER FILING DATE: 1998-10-23  
; EARLIER APPLICATION NUMBER: 08/787,739  
; EARLIER FILING DATE: 1997-01-24  
; EARLIER APPLICATION NUMBER: 08/485,049  
; EARLIER FILING DATE: 1995-06-07  
; EARLIER APPLICATION NUMBER: 08/486,756  
; EARLIER FILING DATE: 1995-06-07  
; EARLIER APPLICATION NUMBER: 08/477,504  
; EARLIER FILING DATE: 1995-06-07  
; EARLIER APPLICATION NUMBER: 08/481,658  
; EARLIER FILING DATE: 1995-06-07  
; EARLIER APPLICATION NUMBER: 08/485,862



```

; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 08/485,863
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 08/487,077
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 08/260,190
; EARLIER FILING DATE: 1994-06-15
; EARLIER APPLICATION NUMBER: 08/177,093
; EARLIER FILING DATE: 1993-12-30
; EARLIER APPLICATION NUMBER: 07/964,589
; EARLIER FILING DATE: 1992-10-21
; EARLIER APPLICATION NUMBER: PV-709-92
; EARLIER FILING DATE: 1992-03-11
; NUMBER OF SEQ ID NOS: 116
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO: 108
; LENGTH: 7
; TYPE: PRT
; ORGANISM: HUMAN
US-09-178-115-108

```

```

Query Match      86.1%; Score 31; DB 4; Length 7;
Best Local Similarity 85.7%; Pred. No. 1.9e+05;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 1 GETRAPL 7
   |||||
Db 1 GETREPL 7

```

## RESULT 6

```

US-09-177-776-108
; Sequence 108, Application US/09177776A
; Patent No. 6297051
; GENERAL INFORMATION:
; APPLICANT: Zavada, Jan
; APPLICANT: Pastorek, Silvia
; TITLE OF INVENTION: MN Gene and Protein
; FILE REFERENCE: D-0021.SA
; CURRENT APPLICATION NUMBER: US/09/177,776A
; EARLIER FILING DATE: 1998-10-23
; EARLIER APPLICATION NUMBER: 08/787,739
; EARLIER FILING DATE: 1997-01-24
; EARLIER APPLICATION NUMBER: 08/485,049
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 08/486,756
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 08/477,504
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 08/481,658
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 08/485,862
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 08/485,863
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 08/487,077
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 08/260,190
; EARLIER FILING DATE: 1994-06-15
; EARLIER APPLICATION NUMBER: 08/177,093
; EARLIER FILING DATE: 1993-12-30
; EARLIER APPLICATION NUMBER: 07/964,589
; EARLIER FILING DATE: 1992-10-21
; EARLIER APPLICATION NUMBER: PV-709-92
; NUMBER OF SEQ ID NOS: 116
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO: 108
; LENGTH: 7
; TYPE: PRT
; ORGANISM: HUMAN
US-09-177-776-108

```

```

Query Match      86.1%; Score 31; DB 4; Length 7;
Best Local Similarity 85.7%; Pred. No. 1.9e+05;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 1 GETRAPL 7
   |||||
Db 1 GETREPL 7

```

## RESULT 7

```

US-09-178-115-109
; Sequence 109, Application US/09178115
; Patent No. 6297041
; GENERAL INFORMATION:
; APPLICANT: Zavada, Jan
; APPLICANT: Pastorek, Silvia
; TITLE OF INVENTION: MN Gene and Protein
; FILE REFERENCE: D-0021.SA
; CURRENT APPLICATION NUMBER: US/09/178,115
; EARLIER FILING DATE: 1998-10-23
; EARLIER APPLICATION NUMBER: 09/177,776
; EARLIER FILING DATE: 1998-10-23
; EARLIER APPLICATION NUMBER: 08/787,739
; EARLIER FILING DATE: 1997-01-24
; EARLIER APPLICATION NUMBER: 08/485,049
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 08/486,756
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 08/477,504
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 08/481,658
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 08/485,862
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 08/485,863
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 08/487,077
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 08/260,190
; EARLIER FILING DATE: 1994-06-15
; EARLIER APPLICATION NUMBER: 08/177,093
; EARLIER FILING DATE: 1993-12-30
; EARLIER APPLICATION NUMBER: 07/964,589
; EARLIER FILING DATE: 1992-10-21
; EARLIER APPLICATION NUMBER: PV-709-92
; NUMBER OF SEQ ID NOS: 116
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO: 109
; LENGTH: 7
; TYPE: PRT
; ORGANISM: HUMAN
US-09-178-115-109

```

```

Query Match      83.3%; Score 30; DB 4; Length 7;
Best Local Similarity 71.4%; Pred. No. 1.9e+05;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 GETRAPL 7
   |||||
Db 1 GETREPL 7

```

## RESULT 8

```

US-09-177-776-109
; Sequence 109, Application US/09177776A
; Patent No. 6297051
; GENERAL INFORMATION:
; APPLICANT: Zavada, Jan
; APPLICANT: Pastorek, Silvia
; APPLICANT: Pastorek, Jaromir

```

TITLE OF INVENTION: MN Gene and Protein  
FILE REFERENCE: D-0021.5A  
CURRENT APPLICATION NUMBER: US/09/177,776A  
CURRENT FILING DATE: 1998-10-23  
EARLIER APPLICATION NUMBER: 08/787,739  
EARLIER FILING DATE: 1997-01-24  
EARLIER APPLICATION NUMBER: 08/485,049  
EARLIER FILING DATE: 1995-06-07  
EARLIER APPLICATION NUMBER: 08/486,756  
EARLIER FILING DATE: 1995-06-07  
EARLIER APPLICATION NUMBER: 08/477,504  
EARLIER FILING DATE: 1995-06-07  
EARLIER APPLICATION NUMBER: 08/481,658  
EARLIER FILING DATE: 1995-06-07  
EARLIER APPLICATION NUMBER: 08/485,862  
EARLIER FILING DATE: 1995-06-07  
EARLIER APPLICATION NUMBER: 08/485,863  
EARLIER FILING DATE: 1995-06-07  
EARLIER APPLICATION NUMBER: 08/487,077  
EARLIER FILING DATE: 1995-06-07  
EARLIER APPLICATION NUMBER: 08/260,190  
EARLIER FILING DATE: 1994-06-15  
EARLIER APPLICATION NUMBER: 08/177,093  
EARLIER FILING DATE: 1993-12-30  
EARLIER APPLICATION NUMBER: 07/964,589  
EARLIER FILING DATE: 1992-10-21  
EARLIER APPLICATION NUMBER: PV-709-92  
EARLIER FILING DATE: 1992-03-11  
NUMBER OF SEQ ID NOS: 116  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 109  
LENGTH: 7  
TYPE: PRT  
ORGANISM: HUMAN  
US-09-177-776-109

Query Match 83.3% Score 30; DB 4; Length 7;  
Best Local Similarity 71.4%; Pred. No. 1.9e+05;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 GETRPL 7  
|:|:|  
Db 1 GOTRPL 7

RESULT 9  
US-08-456-647B-51  
Sequence 51, Application US/08456647B  
Patent No. 5811516  
GENERAL INFORMATION:  
APPLICANT: Lemke Ph.D. et al., Greg E.  
TITLE OF INVENTION: PROTEIN-TYROSINE KINASE GENES  
NUMBER OF SEQUENCES: 54  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 4225 Executive Square, Suite 1400  
CITY: La Jolla  
STATE: CA  
COUNTRY: US  
ZIP: 92037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/456,647B  
FILING DATE: 02-JUN-1995  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/237,401  
FILING DATE: 02-MAY-1994  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/884,486  
FILING DATE: 15-MAY-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Weherelli Ph.D., John R.  
REGISTRATION NUMBER: 31,678  
REFERENCE/DOCKET NUMBER: 07251/007002  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 678-5099  
TELEFAX: (619) 678-5070  
INFORMATION FOR SEQ ID NO: 51:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 54 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-456-647B-51

Query Match 77.8% Score 28; DB 2; Length 54;  
Best Local Similarity 71.4%; Pred. No. 20;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GETRPL 7  
|:|:|  
Db 31 GOTRPL 37

RESULT 10  
US-08-237-401A-51  
Sequence 51, Application US/08237401A  
Patent No. 5837448  
GENERAL INFORMATION:  
APPLICANT: Lemke Ph.D. et al., Greg E.  
TITLE OF INVENTION: PROTEIN-TYROSINE KINASE GENES  
NUMBER OF SEQUENCES: 54  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 4225 Executive Square, Suite 1400  
CITY: La Jolla  
STATE: CA  
COUNTRY: US  
ZIP: 92037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/237,401A  
FILING DATE: 02-MAY-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/884,486  
FILING DATE: 15-MAY-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Halle Ph.D., Lisa A.  
REGISTRATION NUMBER: 38,347  
REFERENCE/DOCKET NUMBER: 07251/007001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 678-5070  
TELEFAX: (619) 678-5099  
INFORMATION FOR SEQ ID NO: 51:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 54 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-237-401A-51

Query Match 77.8% Score 28; DB 2; Length 54;  
Best Local Similarity 71.4%; Pred. No. 20;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GETRAPL 7  
 DB 31 GDTPL 37

RESULT 11  
 US-09-390-326-9  
 ; Sequence 9, Application US/09390326  
 ; Patent No. 6316603

GENERAL INFORMATION:  
 ; APPLICANT: METIGUE, MICHELE A.  
 ; APPLICANT: WICKERSHAM, JOHN A.  
 ; APPLICANT: PINKO, CHRIS  
 ; APPLICANT: SHOMALTER, RICHARD  
 ; APPLICANT: PARAST, CAMRAN V.  
 ; APPLICANT: TEMPICZYK-RUSSEL, ANNA  
 ; APPLICANT: GEHRING, MICHAEL R.  
 ; APPLICANT: MROCKOWSKI, BARBARA  
 ; APPLICANT: KAN, CHEN-CHEN  
 ; APPLICANT: VILLAFRANCA, J. ERNEST  
 ; APPLICANT: APPELT, KRYSZTOF  
 ; TITLE OF INVENTION: MODIFICATIONS OF THE VEGF RECEPTOR-2 PROTEIN AND  
 ; FILE REFERENCE: 0125-0016US  
 ; CURRENT APPLICATION NUMBER: US/09/390,326  
 ; CURRENT FILING DATE: 1999-09-07  
 ; NUMBER OF SEQ ID NOS: 12  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 9  
 ; LENGTH: 367  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-09-390-326-9

Query Match 77.8%; Score 28; DB 4; Length 367;  
 Best Local Similarity 71.4%; Pred. No. 1.6e+02;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GETRAPL 7  
 DB 259 GDTPL 265

RESULT 12  
 US-09-413-814-82  
 ; Sequence 82, Application US/09413814  
 ; Patent No. 6225064  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH  
 ; APPLICANT: Bristol-Myers Squibb, Co.  
 ; APPLICANT: Beyer, Stefan  
 ; APPLICANT: Bioecker, Helmut  
 ; APPLICANT: Brandt, Petra  
 ; APPLICANT: Cino, Paul M  
 ; APPLICANT: Dougherty, Brian A  
 ; APPLICANT: Goldberg, Steven L  
 ; APPLICANT: Hoelle, Gerhard  
 ; APPLICANT: Mueller, Joachim  
 ; APPLICANT: Reichenbach, Hans  
 ; TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polyketide or  
 ; FILE REFERENCE: PCT/US 99/23535  
 ; CURRENT APPLICATION NUMBER: US/09/413,814  
 ; CURRENT FILING DATE: 1999-10-07  
 ; EARLIER APPLICATION NUMBER: DE 198 46 493.2  
 ; EARLIER FILING DATE: 1998-10-09  
 ; NUMBER OF SEQ ID NOS: 107  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 82  
 ; LENGTH: 514  
 ; TYPE: PRT  
 ; ORGANISM: Sorangium cellulosum

US-09-413-814-82

Query Match 77.8%; Score 28; DB 4; Length 514;  
 Best Local Similarity 85.7%; Pred. No. 2.3e+02;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GETRAPL 7  
 DB 45 GLTRAPL 51

RESULT 13  
 US-09-413-814-94  
 ; Sequence 94, Application US/09413814  
 ; Patent No. 6225064  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH  
 ; APPLICANT: Bristol-Myers Squibb, Co.  
 ; APPLICANT: Beyer, Stefan  
 ; APPLICANT: Bioecker, Helmut  
 ; APPLICANT: Brandt, Petra  
 ; APPLICANT: Cino, Paul M  
 ; APPLICANT: Dougherty, Brian A  
 ; APPLICANT: Goldberg, Steven L  
 ; APPLICANT: Hoelle, Gerhard  
 ; APPLICANT: Mueller, Joachim  
 ; APPLICANT: Reichenbach, Hans  
 ; TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polyketide or  
 ; FILE REFERENCE: PCT/US 99/23535  
 ; CURRENT APPLICATION NUMBER: US/09/413,814  
 ; CURRENT FILING DATE: 1999-10-07  
 ; EARLIER APPLICATION NUMBER: DE 198 46 493.2  
 ; EARLIER FILING DATE: 1998-10-09  
 ; NUMBER OF SEQ ID NOS: 107  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 94  
 ; LENGTH: 514  
 ; TYPE: PRT  
 ; ORGANISM: Sorangium cellulosum  
 US-09-413-814-94

Query Match 77.8%; Score 28; DB 4; Length 514;  
 Best Local Similarity 85.7%; Pred. No. 2.3e+02;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GETRAPL 7  
 DB 45 GLTRAPL 51

RESULT 14  
 US-08-796-414B-1  
 ; Sequence 1, Application US/08796414B  
 ; Patent No. 5876987  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Wendy C. Champness, Paul Brian  
 ; APPLICANT: and Todd B. Anderson  
 ; TITLE OF INVENTION: METHOD, DNA AND BACTERIA  
 ; TITLE OF INVENTION: FOR HYPERPRODUCTION OF AN  
 ; TITLE OF INVENTION: ANTIBIOTIC DUE TO  
 ; TITLE OF INVENTION: DISRUPTION OF AN Abba  
 ; NUMBER OF SEQUENCES: 7  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSER: Ian C. McLeod  
 ; STREET: 2190 Commons Parkway  
 ; CITY: Okemos  
 ; STATE: Michigan  
 ; COUNTRY: USA  
 ; ZIP: 48864  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette, 3.5 inch,

```

; COMPUTER: Acer
; OPERATING SYSTEM: MS-DOS 5.00
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/796,414B
; FILING DATE: February 6, 1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Ian C. McLeod
; REGISTRATION NUMBER: 20,931
; REFERENCE/DOCKET NUMBER: MSU 4.1-297
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (517) 347-4100
; TELEFAX: (517) 347-4103
; TELEX: No. 5876987e
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 571
; TYPE: Amino Acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
; MOLECULE TYPE: Amino Acid
; HYPOTHEetical: No
; ANTI-SENSE: No
; FRAGMENT TYPE:
; ORIGINAL SOURCE:
; ORGANISM: Streptomyces coelicolor
; STRAIN: N/A
; INDIVIDUAL ISOLATE:
; DEVELOPMENTAL STAGE:
; HAPOTYPE: N/A
; TISSUE TYPE: N/A
; CELL TYPE: unicellular organism
; CELL LINE:
; ORGANELL:
; IMMEDIATE SOURCE: N/A
; POSITION IN GENOME:
; FEATURE:
; NAME/KEY: Absa1
; LOCATION:
; IDENTIFICATION METHOD: deduced
; OTHER INFORMATION:
; PUBLICATION INFORMATION:
; AUTHORS:
; TITLE:
; JOURNAL:
; VOLUME:
; ISSUE:
; PAGES:
; DATE:
; RELEVANT RESIDUES IN SEQ ID NO:
; US-08-796-414B-1

Query March 77.8%; Score 28; DB 2; Length 571;
Best Local Similarity 95.7%; Pred. No. 2.5e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GETRAPL 7
Db 506 GETRADL 512

```

```

RESULT 15
US-09-121-964-9
; Sequence 9, Application US/09121964
; Patent No. 6124447
; GENERAL INFORMATION:
; APPLICANT: Natco1, Shunji
; TITLE OF INVENTION: NOVEL ENZYME CATALYZING DEPHOSPHORYLATION
; FILE REFERENCE: 32290-144753

```

```

; CURRENT APPLICATION NUMBER: US/09/121,964
; CURRENT FILING DATE: 1998-07-24
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 771
; TYPE: PRT
; ORGANISM: C. elegans
; US-09-121-964-9

Query Match 77.8%; Score 28; DB 3; Length 771;
Best Local Similarity 71.4%; Pred. No. 3.5e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GETRAPL 7
Db 95 GERRAPI 101

Search completed: January 29, 2003, 14:04:17
Job time : 9.61539 secs

```

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2003, 14:00:01 ; Search time 5.38462 Seconds  
(without alignments)  
26.232 Million cell updates/sec

Title: US-09-807-949a-107  
Perfect score: 36  
Sequence: 1 GETRAPL 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 122226 seqs, 20178551 residues

Total number of hits satisfying chosen parameters: 122226

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Published Applications AA.\*  
1: /cgn2\_6/ptodata/1/pubppa/US08\_NEW\_PUB.pep.\*  
2: /cgn2\_6/ptodata/1/pubppa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubppa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubppa/US07\_NEW\_PUB.pep.\*  
5: /cgn2\_6/ptodata/1/pubppa/US07\_PUBCOMB.pep.\*  
6: /cgn2\_6/ptodata/1/pubppa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubppa/US08\_PUBCOMB.pep.\*  
8: /cgn2\_6/ptodata/1/pubppa/US09\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubppa/US09\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/1/pubppa/US10\_NEW\_PUB.pep.\*  
11: /cgn2\_6/ptodata/1/pubppa/US10\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/1/pubppa/US60\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/1/pubppa/US60\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/1/pubppa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	36	100.0	7	10	US-09-947-137-29 Sequence 29, Appl
2	32	88.9	271	10	US-09-925-297-545 Sequence 545, Appl
3	32	88.9	364	9	US-09-764-868-1087 Sequence 1087, Ap
4	32	88.9	612	9	US-09-764-868-663 Sequence 663, Ap
5	32	88.9	820	9	US-09-989-442-118 Sequence 118, Appl
6	30	83.3	865	10	US-09-815-242-10151 Sequence 10151, A
7	30	83.3	867	10	US-09-815-242-11301 Sequence 11301, A
8	30	83.3	868	10	US-09-815-242-5122 Sequence 5122, Ap
9	29	80.6	101	9	US-09-796-692-1971 Sequence 1971, Ap
10	29	80.6	101	9	US-09-796-692-2487 Sequence 2487, Ap
11	29	80.6	198	10	US-09-925-300-1562 Sequence 1562, Ap
12	29	80.6	208	10	US-09-864-761-35865 Sequence 35865, A
13	29	80.6	416	10	US-09-815-242-11301 Sequence 11301, A
14	29	80.6	416	10	US-09-861-752a-1172 Sequence 1172, Appl
15	28	77.8	54	9	US-09-158-722-51 Sequence 51, Appl
16	28	77.8	120	9	US-09-736-457-797 Sequence 797, Appl
17	28	77.8	120	9	US-09-902-941-797 Sequence 797, Appl
18	28	77.8	120	9	US-09-849-626-797 Sequence 797, Appl
19	28	77.8	308	10	US-09-764-864-927 Sequence 927, Appl

20	28	77.8	309	10	US-09-764-864-1366 Sequence 1366, Ap
21	28	77.8	322	10	US-09-942-447-5 Sequence 5, Appl
22	28	77.8	367	9	US-09-939-833-9 Sequence 9, Appl
23	28	77.8	367	10	US-09-939-754-9 Sequence 9, Appl
24	28	77.8	367	10	US-09-939-832-9 Sequence 9, Appl
25	28	77.8	388	10	US-09-925-300-1618 Sequence 1618, Ap
26	28	77.8	427	10	US-09-942-447-4 Sequence 4, Appl
27	28	77.8	618	9	US-09-738-626-3975 Sequence 3975, Ap
28	28	77.8	635	10	US-09-815-242-11209 Sequence 11209, A
29	28	77.8	637	10	US-09-815-242-12058 Sequence 12058, A
30	28	77.8	1073	9	US-09-738-626-4365 Sequence 4365, Ap
31	27	75.0	71	10	US-09-948-080-16 Sequence 16, Appl
32	27	75.0	243	10	US-09-864-761-49091 Sequence 49091, A
33	27	75.0	247	9	US-09-854-133-112 Sequence 112, Appl
34	27	75.0	247	10	US-09-738-973-112 Sequence 112, Appl
35	27	75.0	277	10	US-09-815-242-5708 Sequence 5708, Ap
36	27	75.0	277	10	US-09-815-242-12249 Sequence 12249, A
37	27	75.0	277	10	US-09-815-242-12767 Sequence 12767, A
38	27	75.0	288	10	US-09-815-242-11689 Sequence 11689, A
39	27	75.0	344	10	US-09-815-242-1059 Sequence 1059, Ap
40	27	75.0	369	10	US-09-815-242-10807 Sequence 10807, A
41	27	75.0	408	10	US-09-057-951-4 Sequence 4, Appl
42	27	75.0	408	12	US-10-105-150-4 Sequence 4, Appl
43	27	75.0	430	10	US-09-057-951-2 Sequence 2, Appl
44	27	75.0	430	10	US-09-836-607-2 Sequence 2, Appl
45	27	75.0	430	12	US-10-105-150-2 Sequence 2, Appl

## ALIGNMENTS

```

RESULT 1
US-09-947-137-29
; Sequence 29, Application US/09947137
; Patent No. US20020137023A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Bruce F.
; APPLICANT: Samolova, Tatiana
; TITLE OF INVENTION: Methods and Compositions for Targeting
; TITLE OF INVENTION: Compounds to Muscle
; FILE REFERENCE: 5721-8
; CURRENT APPLICATION NUMBER: US/09/947,137
; CURRENT FILING DATE: 2001-09-05
; PRIOR APPLICATION NUMBER: 09/084,605
; PRIOR FILING DATE: 1998-05-26
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 29
; LENGTH: 7
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Phage display library peptides
US-09-947-137-29

Query Match      100.0%  Score 36;  DB 10;  Length 7;
Best Local Similarity 100.0%;  Pred. No. 1e+05;  0;  Indels 0;
Matches 7;  Conservative 0;  Mismatches 0;

Qy      1 GETRAPL 7
      |||||
Db      1 GETRAPL 7

RESULT 2
US-09-925-297-545
; Sequence 545, Application US/09925297
; Patent No. US20020081659A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA105
; CURRENT APPLICATION NUMBER: US/09/925,297

```

```
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05989
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 928
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 545
; LENGTH: 271
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-925-297-545
```

```
Query Match      88.9%; Score 32; DB 10; Length 271;
Best Local Similarity 85.7%; Pred. No. 11;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
OY 1 GETRAPL 7
    |||||
Db 79 GETRVPL 85
```

```
RESULT 3
US-09-764-868-1087
; Sequence 1087, Application US/09764868
; Patent No. US2002016871A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT232
; CURRENT APPLICATION NUMBER: US/09/764,868
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1087
; LENGTH: 364
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: SITE
; LOCATION: (25)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (81)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-868-1087
```

```
Query Match      88.9%; Score 32; DB 9; Length 364;
Best Local Similarity 85.7%; Pred. No. 15;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
OY 1 GETRAPL 7
    |||||
Db 172 GETRVPL 178
```

```
RESULT 4
US-09-764-868-663
; Sequence 663, Application US/09764868
; Patent No. US2002016871A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT232
; CURRENT APPLICATION NUMBER: US/09/764,868
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 663
; LENGTH: 612
; TYPE: PRT
; ORGANISM: Homo sapiens
```

```
; ORGANISM: Homo sapiens
US-09-764-868-663
```

```
Query Match      88.9%; Score 32; DB 9; Length 612;
Best Local Similarity 85.7%; Pred. No. 26;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
OY 1 GETRAPL 7
    |||||
Db 175 GETRVPL 181
```

```
RESULT 5
US-09-989-442-118
; Sequence 118, Application US/0989442
; Publication No. US20030013649A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: P2308
; CURRENT APPLICATION NUMBER: US/09/989,442
; CURRENT FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: 60/179,065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180,628
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: 60/214,886
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 60/217,487
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,758
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/220,963
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/217,496
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,447
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/218,290
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/225,757
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/226,868
; PRIOR FILING DATE: 2000-08-22
; PRIOR APPLICATION NUMBER: 60/216,647
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: 60/225,267
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/216,880
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: 60/225,270
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/251,869
; PRIOR FILING DATE: 2000-12-08
; PRIOR APPLICATION NUMBER: 60/235,834
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: 60/234,274
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: 60/234,223
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: 60/228,924
; PRIOR FILING DATE: 2000-08-30
; PRIOR APPLICATION NUMBER: 60/224,518
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/236,369
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: 60/224,519
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/220,964
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/241,809
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/249,299
```

/ PRIOR FILING DATE: 2000-11-17  
 / PRIOR APPLICATION NUMBER: 60/236,327  
 / PRIOR FILING DATE: 2000-09-29  
 / PRIOR APPLICATION NUMBER: 60/241,785  
 / PRIOR FILING DATE: 2000-10-20  
 / PRIOR APPLICATION NUMBER: 60/244,617  
 / PRIOR FILING DATE: 2000-11-01  
 / PRIOR APPLICATION NUMBER: 60/225,268  
 / PRIOR FILING DATE: 2000-08-14  
 / PRIOR APPLICATION NUMBER: 60/236,368  
 / PRIOR FILING DATE: 2000-09-29  
 / PRIOR APPLICATION NUMBER: 60/251,856  
 / PRIOR FILING DATE: 2000-12-08  
 / PRIOR APPLICATION NUMBER: 60/251,868  
 / PRIOR FILING DATE: 2000-12-08  
 / PRIOR APPLICATION NUMBER: 60/229,344  
 / PRIOR FILING DATE: 2000-09-01  
 / PRIOR APPLICATION NUMBER: 60/234,997  
 / PRIOR FILING DATE: 2000-09-25  
 / PRIOR APPLICATION NUMBER: 60/229,343  
 / PRIOR FILING DATE: 2000-09-01  
 / PRIOR APPLICATION NUMBER: 60/229,345  
 / PRIOR FILING DATE: 2000-09-01  
 / PRIOR APPLICATION NUMBER: 60/229,287  
 / PRIOR FILING DATE: 2000-09-01  
 / PRIOR APPLICATION NUMBER: 60/229,513  
 / PRIOR FILING DATE: 2000-09-05  
 / PRIOR APPLICATION NUMBER: 60/231,413  
 / PRIOR FILING DATE: 2000-09-08  
 / PRIOR APPLICATION NUMBER: 60/229,509  
 / PRIOR FILING DATE: 2000-09-05  
 / PRIOR APPLICATION NUMBER: 60/236,367  
 / PRIOR FILING DATE: 2000-09-29  
 / PRIOR APPLICATION NUMBER: 60/237,039  
 / PRIOR FILING DATE: 2000-10-02  
 / PRIOR APPLICATION NUMBER: 60/237,038  
 / PRIOR FILING DATE: 2000-10-02  
 / PRIOR APPLICATION NUMBER: 60/236,370  
 / PRIOR FILING DATE: 2000-09-29  
 / PRIOR APPLICATION NUMBER: 60/236,802  
 / PRIOR FILING DATE: 2000-10-02  
 / PRIOR APPLICATION NUMBER: 60/237,037  
 / PRIOR FILING DATE: 2000-10-02  
 / PRIOR APPLICATION NUMBER: 60/237,040  
 / PRIOR FILING DATE: 2000-10-02  
 / PRIOR APPLICATION NUMBER: 60/240,960  
 / PRIOR FILING DATE: 2000-10-20  
 / PRIOR APPLICATION NUMBER: 60/239,935  
 / PRIOR FILING DATE: 2000-10-13  
 / PRIOR APPLICATION NUMBER: 60/239,937  
 / PRIOR FILING DATE: 2000-10-13  
 / PRIOR APPLICATION NUMBER: 60/241,787  
 / PRIOR FILING DATE: 2000-10-20  
 / PRIOR APPLICATION NUMBER: 60/246,474  
 / PRIOR FILING DATE: 2000-11-08  
 / PRIOR APPLICATION NUMBER: 60/246,532  
 / PRIOR FILING DATE: 2000-11-08  
 / PRIOR APPLICATION NUMBER: 60/249,216  
 / PRIOR FILING DATE: 2000-11-17  
 / PRIOR APPLICATION NUMBER: 60/249,210  
 / PRIOR FILING DATE: 2000-11-17  
 / PRIOR APPLICATION NUMBER: 60/226,681  
 / PRIOR FILING DATE: 2000-08-22  
 / PRIOR APPLICATION NUMBER: 60/225,759  
 / PRIOR FILING DATE: 2000-08-14  
 / PRIOR APPLICATION NUMBER: 60/225,213  
 / PRIOR FILING DATE: 2000-08-14  
 / PRIOR APPLICATION NUMBER: 60/227,182  
 / PRIOR FILING DATE: 2000-08-22  
 / PRIOR APPLICATION NUMBER: 60/225,214  
 / PRIOR FILING DATE: 2000-08-14  
 / PRIOR APPLICATION NUMBER: 60/235,836  
 / PRIOR FILING DATE: 2000-09-27

/ PRIOR APPLICATION NUMBER: 60/230,438  
 / PRIOR FILING DATE: 2000-09-06  
 / PRIOR APPLICATION NUMBER: 60/215,135  
 / PRIOR FILING DATE: 2000-06-30  
 / PRIOR APPLICATION NUMBER: 60/225,266  
 / PRIOR FILING DATE: 2000-08-14  
 / PRIOR APPLICATION NUMBER: 60/249,218  
 / PRIOR FILING DATE: 2000-11-17  
 / PRIOR APPLICATION NUMBER: 60/249,208  
 / PRIOR FILING DATE: 2000-11-17  
 / PRIOR APPLICATION NUMBER: 60/249,213  
 / PRIOR FILING DATE: 2000-11-17  
 / PRIOR APPLICATION NUMBER: 60/249,212  
 / PRIOR FILING DATE: 2000-11-17  
 / PRIOR APPLICATION NUMBER: 60/249,207  
 / PRIOR FILING DATE: 2000-11-17  
 / PRIOR APPLICATION NUMBER: 60/249,245  
 / PRIOR FILING DATE: 2000-11-17  
 / PRIOR APPLICATION NUMBER: 60/249,244  
 / PRIOR FILING DATE: 2000-11-17  
 / PRIOR APPLICATION NUMBER: 60/249,217  
 / PRIOR FILING DATE: 2000-11-17  
 / PRIOR APPLICATION NUMBER: 60/249,211  
 / PRIOR FILING DATE: 2000-11-17  
 / PRIOR APPLICATION NUMBER: 60/249,215  
 / PRIOR FILING DATE: 2000-11-17  
 / PRIOR APPLICATION NUMBER: 60/249,264  
 / PRIOR FILING DATE: 2000-11-17  
 / PRIOR APPLICATION NUMBER: 60/249,214  
 / PRIOR FILING DATE: 2000-11-17  
 / PRIOR APPLICATION NUMBER: 60/249,297  
 / PRIOR FILING DATE: 2000-11-17  
 / PRIOR APPLICATION NUMBER: 60/232,400  
 / PRIOR FILING DATE: 2000-09-14  
 / PRIOR APPLICATION NUMBER: 60/231,242  
 / PRIOR FILING DATE: 2000-09-08  
 / PRIOR APPLICATION NUMBER: 60/232,081  
 / PRIOR FILING DATE: 2000-09-08  
 / PRIOR APPLICATION NUMBER: 60/232,080  
 / PRIOR FILING DATE: 2000-09-08  
 / PRIOR APPLICATION NUMBER: 60/231,414  
 / PRIOR FILING DATE: 2000-09-08  
 / PRIOR APPLICATION NUMBER: 60/231,244  
 / PRIOR FILING DATE: 2000-09-08  
 / PRIOR APPLICATION NUMBER: 60/233,064  
 / PRIOR FILING DATE: 2000-09-14  
 / PRIOR APPLICATION NUMBER: 60/233,063  
 / PRIOR FILING DATE: 2000-09-14  
 / PRIOR APPLICATION NUMBER: 60/232,397  
 / PRIOR FILING DATE: 2000-09-14  
 / PRIOR APPLICATION NUMBER: 60/232,399  
 / PRIOR FILING DATE: 2000-09-14  
 / PRIOR APPLICATION NUMBER: 60/232,401  
 / PRIOR FILING DATE: 2000-09-14  
 / PRIOR APPLICATION NUMBER: 60/241,808  
 / PRIOR FILING DATE: 2000-10-20  
 / PRIOR APPLICATION NUMBER: 60/241,826  
 / PRIOR FILING DATE: 2000-10-20  
 / PRIOR APPLICATION NUMBER: 60/241,786  
 / PRIOR FILING DATE: 2000-10-20  
 / PRIOR APPLICATION NUMBER: 60/241,221  
 / PRIOR FILING DATE: 2000-10-20  
 / PRIOR APPLICATION NUMBER: 60/246,475  
 / PRIOR FILING DATE: 2000-11-08  
 / PRIOR APPLICATION NUMBER: 60/231,243  
 / PRIOR FILING DATE: 2000-09-08  
 / PRIOR APPLICATION NUMBER: 60/233,065  
 / PRIOR FILING DATE: 2000-09-14

Query Match 88.9%; Score 32; DB 9; Length 820;  
 Best Local Similarity 85.7%; Pred. No. 36;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GETRAPL 7  
Db 172 GETRAPL 178

## RESULT 6

US-09-815-242-10151  
Sequence 10151, Application US/09815242  
Patent No. US20020061569A1  
GENERAL INFORMATION:  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Karl L.  
APPLICANT: Zyekind, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard  
TITLE OF INVENTION: Identification of Essential Genes in  
FILE REFERENCE: ELITRA.011A  
CURRENT APPLICATION NUMBER: US/09/815,242  
PRIOR FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 14110  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 10151  
LENGTH: 865  
TYPE: PRT  
ORGANISM: Escherichia coli  
US-09-815-242-10151

Query Match 83.3%; Score 30; DB 10; Length 865;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 ETRAPL 7  
Db 791 ETRAPL 796

RESULT 7  
US-09-815-242-13848  
Sequence 13848, Application US/09815242  
Patent No. US20020061569A1  
GENERAL INFORMATION:  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Karl L.  
APPLICANT: Zyekind, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard  
TITLE OF INVENTION: Identification of Essential Genes in  
FILE REFERENCE: ELITRA.011A  
CURRENT APPLICATION NUMBER: US/09/815,242  
PRIOR FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 14110  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 13848  
LENGTH: 867  
TYPE: PRT  
ORGANISM: Salmonella typhi  
FEATURE:  
NAME/KEY: VARIANT  
LOCATION: (1)...(867)  
OTHER INFORMATION: Xaa = Any Amino Acid  
US-09-815-242-13848

Query Match 83.3%; Score 30; DB 10; Length 867;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 ETRAPL 7  
Db 793 ETRAPL 798

RESULT 8  
US-09-815-242-5122  
Sequence 5122, Application US/09815242  
Patent No. US20020061569A1  
GENERAL INFORMATION:  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Karl L.  
APPLICANT: Zyekind, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard  
TITLE OF INVENTION: Identification of Essential Genes in  
FILE REFERENCE: ELITRA.011A  
CURRENT APPLICATION NUMBER: US/09/815,242  
CURRENT FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 14110  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 5122  
LENGTH: 868  
TYPE: PRT  
ORGANISM: Pseudomonas aeruginosa  
US-09-815-242-5122



Query Match 83.3%; Score 30; DB 10; Length 868;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 ETRAPL 7  
|:|||||  
Db 793 ETRAPL 798

## RESULT 9

US-09-796-692-1971  
; Sequence 1971, Application US/09796692  
; Publication No. US20020198362A1  
; GENERAL INFORMATION:  
; APPLICANT: Gaiger, Alexander  
; APPLICANT: Mannion, Paul A.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY  
; FILE REFERENCE: 2077.001200  
; CURRENT APPLICATION NUMBER: US/09/796,692  
; CURRENT FILING DATE: 2001-03-01  
; PRIOR APPLICATION NUMBER: 60/186,126  
; PRIOR FILING DATE: 2000-03-01  
; PRIOR APPLICATION NUMBER: 60/190,479  
; PRIOR FILING DATE: 2000-03-17  
; PRIOR APPLICATION NUMBER: 60/200,545  
; PRIOR FILING DATE: 2000-04-27  
; PRIOR APPLICATION NUMBER: 60/200,303  
; PRIOR FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: 60/200,779  
; PRIOR FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: 60/200,999  
; PRIOR FILING DATE: 2000-05-01  
; PRIOR APPLICATION NUMBER: 60/202,084  
; PRIOR FILING DATE: 2000-05-04  
; PRIOR APPLICATION NUMBER: 60/206,201  
; PRIOR FILING DATE: 2000-05-22  
; PRIOR APPLICATION NUMBER: 60/218,950  
; PRIOR FILING DATE: 2000-07-14  
; PRIOR APPLICATION NUMBER: 60/222,903  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: 60/223,416  
; PRIOR FILING DATE: 2000-08-04  
; PRIOR APPLICATION NUMBER: 60/223,378  
; PRIOR FILING DATE: 2000-08-07  
; NUMBER OF SEQ ID NOS: 9597  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 1971  
; LENGTH: 101  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-796-692-1971

Query Match 80.6%; Score 29; DB 9; Length 101;  
Best Local Similarity 83.3%; Pred. No. 18;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GETRAP 6  
|:|||||  
Db 74 GETRAP 79

## RESULT 10

US-09-796-692-2487  
; Sequence 2487, Application US/09796692  
; Publication No. US20020198362A1  
; GENERAL INFORMATION:  
; APPLICANT: Gaiger, Alexander  
; APPLICANT: Algate, Paul A.  
; APPLICANT: Mannion, Jane  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY

; TITLE OF INVENTION: HEMATOLOGICAL MALIGNANCIES  
; FILE REFERENCE: 2077.001200  
; CURRENT APPLICATION NUMBER: US/09/796,692  
; CURRENT FILING DATE: 2001-03-01  
; PRIOR APPLICATION NUMBER: 60/186,126  
; PRIOR FILING DATE: 2000-03-01  
; PRIOR APPLICATION NUMBER: 60/190,479  
; PRIOR FILING DATE: 2000-03-17  
; PRIOR APPLICATION NUMBER: 60/200,545  
; PRIOR FILING DATE: 2000-04-27  
; PRIOR APPLICATION NUMBER: 60/200,303  
; PRIOR FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: 60/200,779  
; PRIOR FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: 60/200,999  
; PRIOR FILING DATE: 2000-05-01  
; PRIOR APPLICATION NUMBER: 60/202,084  
; PRIOR FILING DATE: 2000-05-04  
; PRIOR APPLICATION NUMBER: 60/206,201  
; PRIOR FILING DATE: 2000-05-22  
; PRIOR APPLICATION NUMBER: 60/218,950  
; PRIOR FILING DATE: 2000-07-14  
; PRIOR APPLICATION NUMBER: 60/222,903  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: 60/223,416  
; PRIOR FILING DATE: 2000-08-04  
; PRIOR APPLICATION NUMBER: 60/223,378  
; PRIOR FILING DATE: 2000-08-07  
; NUMBER OF SEQ ID NOS: 9597  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 2487  
; LENGTH: 101  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-796-692-2487

Query Match 80.6%; Score 29; DB 9; Length 101;  
Best Local Similarity 83.3%; Pred. No. 18;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GETRAP 6  
|:|||||  
Db 74 GETRAP 79

RESULT 11  
US-09-925-300-1562  
; Sequence 1562, Application US/09925300  
; Patent No. US20020151681A1  
; GENERAL INFORMATION:  
; APPLICANT: Craig Roegen,  
; APPLICANT: Steve Ruben,  
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
; FILE REFERENCE: PA101  
; CURRENT APPLICATION NUMBER: US/09/925,300  
; PRIOR FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: PCT/US00/05988  
; PRIOR FILING DATE: 2000-03-08  
; PRIOR APPLICATION NUMBER: 60/124,270  
; NUMBER OF SEQ ID NOS: 1890  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1562  
; LENGTH: 198  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (4)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (193)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

US-09-925-300-1562

Query Match 80.6%; Score 29; DB 10; Length 198;  
Best Local Similarity 71.4%; Pred. No. 35;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GETRAP 7  
|||:|  
Db 55 GETRAPV 61

RESULT 12

US-09-864-761-35865  
; Sequence 35865, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharon G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wenheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY  
; FILE REFERENCE: Aeomica-X-1  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263,6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,355  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 09/774,203  
; PRIOR FILING DATE: 2001-01-29  
; NUMBER OF SEQ ID NOS: 49117  
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 35865  
; LENGTH: 208  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
OTHER INFORMATION: MAP TO AP00504.1  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.3  
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.99  
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1  
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.4

OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1  
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.3  
OTHER INFORMATION: EXPRESSED IN HEL100, SIGNAL = 1.4  
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.2  
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1  
OTHER INFORMATION: EST\_HUMAN HIT: BE795445.1, EVALUO 9.00e-64  
OTHER INFORMATION: SWISSPROT HIT: Q13625, EVALUO 1.70e-02  
US-09-864-761-35865

Query Match 80.6%; Score 29; DB 10; Length 208;  
Best Local Similarity 83.3%; Pred. No. 37;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GETRAP 6  
|||:|  
Db 201 GETRSP 206

RESULT 13

US-09-815-242-11301  
; Sequence 11301, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Karl L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; TITLE OF INVENTION: Prokaryotes  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 11301  
; LENGTH: 416  
; TYPE: PRT  
; ORGANISM: Helicobacter pylori  
US-09-815-242-11301

Query Match 80.6%; Score 29; DB 10; Length 416;  
Best Local Similarity 83.3%; Pred. No. 77;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GETRAP 6  
|||:|  
Db 346 GETRSP 351

RESULT 14

US-09-881-752A-172  
; Sequence 172, Application US/09881752A  
; Patent No. US20020115078A1  
; GENERAL INFORMATION:  
; APPLICANT: Kleantous, Harold

```

; APPLICANT: Al-Garawi, Amal
; APPLICANT: Miller, Charles
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Oomen, Raymond P.
; TITLE OF INVENTION: Identification of polynucleotides
; TITLE OF INVENTION: Encoding No. US20020150784Helicobacter Polypeptides in the
; FILE REFERENCE: 06132/041002
; CURRENT APPLICATION NUMBER: US/09/881,752A
; CURRENT FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 08/833,457
; PRIOR FILING DATE: 1997-04-01
; NUMBER OF SEQ ID NOS: 370
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 172
; LENGTH: 416
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-09-881-752A-172

```

```

Query Match
Best Local Similarity 80.6%; Score 29; DB 10; Length 416;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 GETRAP 6
Db 346 GETRSP 351

```

```

RESULT 15
US-09-158-722-51
; Sequence 51, Application US/09158722
; Publication No. US2003013848A1
; GENERAL INFORMATION:
; APPLICANT: Lemke Ph.D. et al., Greg B.
; TITLE OF INVENTION: PROTEIN-TYROSINE KINASE GENES
; NUMBER OF SEQUENCES: 54
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: US
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/158,722
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/456,647
; FILING DATE: 02-JUN-1995
; APPLICATION NUMBER: US 08/237,401
; FILING DATE: 02-MAY-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/884,486
; FILING DATE: 15-MAY-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Weherrell Ph.D., John R.
; REGISTRATION NUMBER: 31,678
; REFERENCE/DOCKET NUMBER: 07251/007002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 678-5070
; TELEFAX: (619) 678-5099
; INFORMATION FOR SEQ ID NO: 51:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 54 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
;

```

```

; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-158-722-51

```

```

Query Match
Best Local Similarity 77.8%; Score 28; DB 9; Length 54;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 1 GETRAPL 7
Db 31 GDTLPL 37

```

```

Search completed: January 29, 2003, 14:04:54
Job time : 6.38462 secs

```

**THIS PAGE BLANK (USPTO)**



PT A molecule which specifically binds to a site on MN protein  
PT (oncoprotein) and prevents adhesion of vertebrate cells to the protein,  
PT useful for treating preneoplastic or neoplastic diseases such as cancer  
PT

PS Claim 5; Page 71; 154pp; English.

XX The invention relates to the inhibition of cell adhesion mediated by  
CC the MN oncoprotein (also known as the MN/CA IX isoenzyme or the MN/G250  
CC protein). The MN protein is a tumour-associated adhesion molecule which  
CC comprises a proteoglycan-like (PG) domain (AAB03017) which contains the  
CC protein's binding site, and a carbonic anhydrase (CA) domain (AAB03018).  
CC Abnormal expression of the MN protein is associated with tumorigenicity.  
CC The invention encompasses molecules (e.g., proteins and peptides) which  
CC which specifically bind to a site on the MN protein, thereby preventing  
CC adhesion of vertebrate cells to the protein in a cell adhesion assay. It  
CC also encompasses MN proteins or MN protein fragments which can be added  
CC to the extracellular environment to prevent the adhesion of vertebrate  
CC cells to each other. The invention also relates to the identification of  
CC the binding site of the MN protein and to a method of identifying a site  
CC on an MN protein to which cells adhere, comprising testing a series of  
CC overlapping peptides from the protein in a cell adhesion assay. The  
CC invention encompasses a vector comprising an expression control sequence  
CC operatively linked to a nucleic acid encoding the variable domain of a  
CC MN-specific antibody, where the domains are separated by a flexible  
CC linker peptide (AAB03035) and the vector inhibits the growth of a  
CC vertebrate preneoplastic or neoplastic cell that abnormally expresses MN  
CC protein. The invention also encompasses a vector comprising a  
CC nucleic acid encoding a cytotoxic protein or peptide operatively linked  
CC to the MN gene promoter, which inhibits the growth of a vertebrate  
CC preneoplastic or neoplastic cell. Also claimed is a repressor complex  
CC that binds to the MN gene promoter (AAB52473). MN proteins and peptides,  
CC MN-binding proteins and peptides, and expression vectors encoding such  
CC proteins and peptides are useful for treating patients with  
CC preneoplastic or neoplastic disease (e.g., cancers) associated with or  
CC characterised by abnormal MN expression. Sequences AAB03032-B03034 and  
CC AAB03055-B03058 represent synthetic phage display library peptides which  
CC bind to the CA domain of the human MN protein (AAB03005).  
XX

SO Sequence 7 AA;

Query Match 100.0%; Score 37; DB 21; Length 7;  
Best Local Similarity 100.0%; Pred. No. 7.8e+05;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GETREP 7  
Db 1 GETREP 7

RESULT 2

AA018695  
ID AA018695 standard; Protein; 243 AA.

XX AA018695;

DT 12-OCT-2001 (first entry)

DB Peptide #5129 encoded by probe for measuring cervical gene expression.

XX Probe; human; microarray; gene expression; cervical epithelial cell;  
KM cervical cancer.

XX Homo sapiens.

XX WO200157278-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00670.

XX 04-FEB-2000; 2000US-0180312.  
PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488901/53.

PT Human genome-derived single exon nucleic acid probes useful for  
PT analyzing gene expression in human cervical epithelial cells -

PS Claim 27; SEQ ID NO 23521; 487pp; English.

XX The present invention relates to human single exon nucleic acid probes  
CC (SNP: see A0110068-A0128459). The present sequence is a peptide encoded  
CC by one such probe. The SNPs are derived from human HeLa cells. The SNPs  
CC can be used to produce a single exon microarray, which can be used for  
CC measuring human gene expression in a sample derived from human cervical  
CC epithelial cells. By measuring gene expression, the probes are therefore  
CC useful in grading and/or staging of diseases of the cervix, notably  
CC cervical cancer.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

SO Sequence 243 AA;

Query Match 89.2%; Score 33; DB 22; Length 243;  
Best Local Similarity 100.0%; Pred. No. 61;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GETREP 6  
Db 41 GETREP 46

RESULT 3

AAB96287  
ID AAB96287 standard; Protein; 613 AA.

XX AAB96287;

DT 29-OCT-2001 (first entry)

DB Putative P. abyssi ferredoxin oxidoreductase #6.

XX Hyperthermophilic archaeon; hyperthermophilic protein.

XX Pyrococcus abyssi.

XX FR2792651-A1.

PD 27-OCT-2000.

PF 21-APR-1999; 99FR-0005034.

XX 21-APR-1999; 99FR-0005034.

XX (CNRS ) CNRS CENT NAT RECH SCI.  
PA (IFRE-) IFREMER INST FR RECH EXPL MER.

XX Forterre P, Thierry JC, Prieur D, Dietrich J, Lecompte O;

PI Querellou J, Weissbach J, Saurin W, Heilig R;

XX WPI; 2001-126236/14.

PT New nucleotide sequences isolated from Pyrococcus abyssi encode  
PT proteins useful in industry -

PS Claim 7; Pages 946-948; 1657pp; French.

XX

CC The present invention relates to the genomic sequence of *Pyrococcus*

CC abyssal (see AAF66431 and AAH41233-7) and *P. abyssal* proteins. *P. abyssal* is

CC a hyperthermophilic archaeon, which is isolated from deep-sea

CC hydrothermal vents. The present sequence is one such *P. abyssal* protein.

CC The proteins of the present invention have various potential industrial

CC uses, since the proteins are stable at very high temperatures, some up to

CC 110 degrees centigrade.

CC Note: This patent is in the same patent family as WO200065062, which

CC contains additional sequences as shown in AAB99132-AAB99143,

CC AAH75903-AAH75920 and AAG66436.

CC

XX

SO Sequence 613 AA;

QY 1 GETREP 6 89.2%; Score 33; DB 22; Length 613;

Db 547 GETREP 552

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 4

ID AAU91287 standard; Protein; 804 AA.

XX

XX AAU91287;

XX

XX 18-JUN-2002 (first entry)

XX

DE Human NOV5f protein.

XX

KW Human; NOVX; gene therapy; cardiomyopathy; atherosclerosis;

KW diabetes; cell signal processing; metabolic pathway modulation;

KW inflammation; autoimmune disorder; scleroderma; transplantation;

KW allergy; systemic lupus erythematosus; haemophilia; Alzheimer's disease;

KW graft versus host disease; Lesch-Nyhan syndrome; periodontitis;

KW pancreatitis; musculoskeletal disorder; Parkinson's disease;

KW Huntington's disease; behavioural disorder; pain; obesity; wound healing;

KW neurodegenerative disorder; neuropsychiatric disorder; hypertension;

KW growth disorder; reproductive disorder; lung disease.

XX

OS Homo sapiens.

XX

XX WO200216600-A2.

XX

XX 28-FEB-2002.

XX

XX 27-AUG-2001; 2001WO-US26518.

XX

XX 25-AUG-2000; 2000US-227800P.

XX 25-AUG-2000; 2000US-228205P.

XX 25-AUG-2000; 2000US-228324P.

XX 30-AUG-2000; 2000US-228997P.

XX 30-AUG-2000; 2000US-229185P.

XX 01-SEP-2000; 2000US-229780P.

XX 01-SEP-2000; 2000US-229848P.

XX 01-SEP-2000; 2000US-229850P.

XX 22-JAN-2001; 2001US-26337P.

XX 31-JAN-2001; 2001US-265518P.

XX 15-MAR-2001; 2001US-276451P.

XX 27-MAR-2001; 2001US-279196P.

XX 24-AUG-2001; 2001US-0393398.

XX

PA (CURA-) CURAGEN CORP.

XX

XX Gerlach V, MacDougall JR, Smithson G, Stone DJ, Ellerman K,

PI Spletter KA, Zernusson BD, Raetelli L, Verney CAM, Patucrajan M,

PI Tchervet VT, Padigaru M, Taupier RJ;

XX

DR WPI; 2002-292064/33.

DR N-PSDB; ABX55571.

XX

XX New isolated cytoplasmic, nuclear, membrane bound and secreted

PT polypeptides, termed NOVX, useful for treating inflammation, autoimmune

PT disorders, haemophilia, Lesch-Nyhan syndrome, pancreatitis,

PT musculoskeletal disorders

XX

XX

PS Claim 1; Page 69; 245pp; English.

XX

CC The invention relates to an isolated cytoplasmic, nuclear, membrane bound

CC or secreted polypeptide, designated NOVX (actually NOV1, 2a, 2b, 3a, 3b,

CC 4, 5a, 5b, 5c, 5d, 5e, 5f, 5g, 5h, 5i, 6, 7 and 8), a variant of NOVX, a

CC mature form, or a variant of the mature form of NOVX. Also included

CC are a polynucleotide encoding NOVX (or its complement), a vector

CC comprising the polynucleotide, a cell comprising the vector, an

CC anti-NOVX antibody, determining the presence of NOVX in a sample

CC using the antibody, determining the presence of NOVX polynucleotide in

CC a sample using a probe which binds to NOVX polynucleotide, identifying a

CC an agent which binds to NOVX (including modulators of NOVX). NOVX, the

CC polynucleotide and the antibody are useful for diagnosing, treating

CC or preventing a NOVX-associated disorder selected from cardiomyopathy,

CC atherosclerosis, diabetes, a disorder related to cell signal processing

CC and metabolic pathway modulation, inflammation, autoimmune disorders,

CC scleroderma, transplantation, allergies, systemic lupus erythematosus,

CC haemophilia, graft versus host disease, Alzheimer's disease, stroke,

CC Lesch-Nyhan syndrome, periodontitis, pancreatitis, musculoskeletal

CC disorders, Parkinson's disease, Huntington's disease, behavioural

CC disorders, pain, neurodegenerative and neuropsychiatric disorders,

CC hypertension, wound healing, obesity, growth and reproductive

CC disorders, lung diseases and many other diseases and disorders listed in

CC the specification. NOVX, the polynucleotide and the antibody are useful

CC in screening assays, detection assays (e.g., chromosomal mapping, tissue

CC typing, forensic biology), predictive medicine (e.g., diagnostic assays,

CC prognostic assays, monitoring clinical trials and pharmacogenomic), and

CC in methods of treatment (e.g., therapeutic and prophylactic). NOVX is

CC useful as immunogen to produce antibodies immunospecific for NOVX, as

CC vaccines to screen for potential agonist and antagonist compounds, and

CC as bait protein in a two-hybrid or three-hybrid assay. The polynucleotide

CC is useful in gene therapy, to express NOVX, to detect NOVX mRNA

CC or a genetic lesion in a NOVX gene, and to modulate NOVX activity. The

CC vector is useful for producing non-human transgenic animals. The antibody

CC is useful for isolating, and purifying NOVX and to monitor protein levels

CC in tissue as part of a clinical testing procedure. The present sequence

CC represents a NOVX protein.

XX

SO Sequence 804 AA;

QY 1 GETREP 6 89.2%; Score 33; DB 23; Length 804;

Db 749 GETREP 754

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 5

ID AAU91284 standard; Protein; 825 AA.

XX

XX AAU91284;

XX

XX 18-JUN-2002 (first entry)

XX

DE Human NOV5c protein.

XX

XX Human; NOVX; gene therapy; cardiomyopathy; atherosclerosis;

KW diabetes; cell signal processing; metabolic pathway modulation;

KW inflammation; autoimmune disorder; scleroderma; transplantation;

KW allergy; systemic lupus erythematosus; haemophilia; Alzheimer's disease;

KW graft versus host disease; Lesch-Nyhan syndrome; periodontitis;

KW pancreatitis; musculoskeletal disorder; Parkinson's disease;

KW Huntington's disease; behavioural disorder; pain; obesity; wound healing;

KM neurodegenerative disorder; neuropsychiatric disorder; hypertension;  
 KM growth disorder; reproductive disorder; lung disease.  
 XX  
 OS Homo sapiens.  
 PN WO200216600-A2.  
 PD 28-FEB-2002.  
 XX  
 PD 27-AUG-2001; 2001WO-US26518.  
 XX  
 XX 25-AUG-2000; 2000US-227800P.  
 PR 25-AUG-2000; 2000US-228205P.  
 PR 25-AUG-2000; 2000US-228324P.  
 PR 30-AUG-2000; 2000US-228979P.  
 PR 30-AUG-2000; 2000US-229185P.  
 PR 01-SEP-2000; 2000US-229780P.  
 PR 01-SEP-2000; 2000US-229848P.  
 PR 01-SEP-2000; 2000US-229850P.  
 PR 22-JAN-2001; 2001US-263337P.  
 PR 31-JAN-2001; 2001US-265518P.  
 PR 15-MAR-2001; 2001US-276451P.  
 PR 27-MAR-2001; 2001US-279196P.  
 PR 24-AUG-2001; 2001US-0393398.  
 XX  
 XX (CURA-) CURAGEN CORP.  
 PA Gerlach V, MacDougall JR, Smithson G, Stone DJ, Ellerman K;  
 PI Sytsek KA, Zernhusen BD, Rastelli L, Verney CAM, Paturajan M;  
 PI Tchernev VT, Padigar M, Taupier RJ;  
 XX  
 XX WPI; 2002-292064/33.  
 DR N-PSDB; ABK55568.  
 XX  
 PT New isolated cytoplasmic, nuclear, membrane bound and secreted  
 PT polypeptides, termed NOVX, useful for treating inflammation, autoimmune  
 PT disorders, haemophilia, Lesch-Nyhan syndrome, pancreatitis,  
 PT musculoskeletal disorders -  
 XX  
 PS Claim 1; Page 60; 245pp; English.  
 XX  
 CC The invention relates to an isolated cytoplasmic, nuclear, membrane bound  
 CC or secreted polypeptide, designated NOVX (actually NOV1, 2a, 2b, 3a, 3b,  
 CC 4, 5a, 5b, 5c, 5d, 5e, 5f, 5g, 5h, 5i, 6, 7 and 8), a variant of NOVX, a  
 CC mature form, or a variant of the mature form of NOVX. Also included  
 CC are a polynucleotide encoding NOVX (or its complement), a vector  
 CC comprising the polynucleotide, a cell comprising the vector, an  
 CC anti-NOVX antibody, determining the presence of NOVX polynucleotide in  
 CC a sample using a probe which binds to NOVX polynucleotide, identifying a  
 CC an agent which binds to NOVX (including modulators of NOVX), NOVX, the  
 CC polynucleotide and the antibody are useful for diagnosing, treating  
 CC or preventing a NOVX-associated disorder selected from cardiomyopathy,  
 CC atherosclerosis, diabetes, a disorder related to cell signal processing  
 CC and metabolic pathway modulation, inflammation, autoimmune disorders,  
 CC scleroderma, transplantation, allergies, systemic lupus erythematosus,  
 CC haemophilia, graft versus host disease, Alzheimer's disease, stroke,  
 CC Lesch-Nyhan syndrome, periodontitis, pancreatitis, musculoskeletal  
 CC disorders, Parkinson's disease, Huntington's disease, behavioural  
 CC disorders, pain, neurodegenerative and neuropsychiatric disorders,  
 CC hypertension, wound healing, obesity, growth and reproductive  
 CC disorders, lung diseases and many other diseases and disorders listed in  
 CC the specification. NOVX, the polynucleotide and the antibody are useful  
 CC in screening assays, detection assays (e.g., chromosomal mapping, tissue  
 CC typing, forensic biology), predictive medicine (e.g., diagnostic assays,  
 CC prognostic assays, monitoring clinical trials and pharmacogenomic), and  
 CC in methods of treatment (e.g., therapeutic and prophylactic). NOVX is  
 CC useful as immunogen to produce antibodies immunospecific for NOVX, as  
 CC vaccines to screen for potential agonist and antagonist compounds, and  
 CC as bait protein in a two-hybrid or three-hybrid assay. The polynucleotide  
 CC is useful in gene therapy, to express NOVX, to detect NOVX mRNA  
 CC or a genetic lesion in a NOVX gene, and to modulate NOVX activity. The  
 CC vector is useful for producing non-human transgenic animals. The antibody

CC is useful for isolating, and purifying NOVX and to monitor protein levels  
 CC in tissue as part of a clinical testing procedure. The present sequence  
 CC represents a NOVX protein.  
 XX  
 SO Sequence 825 AA;  
 Query Match 89.2%; Score 33; DB 23; Length 825;  
 Best Local Similarity 100.0%; Pred. No. 2.3e-02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 DY 1 GETREP 6  
 DB 769 GETREP 774  
 RESULT 6  
 ID AAU91289 standard; Protein; 847 AA.  
 XX  
 AC AAU91289;  
 XX  
 DT 18-JUN-2002 (first entry)  
 XX  
 DE Human NOV5h protein.  
 XX  
 KW Human; NOVX; gene therapy; cardiomyopathy; atherosclerosis;  
 KW diabetes; cell signal processing; metabolic pathway modulation;  
 KW inflammation; autoimmune disorder; scleroderma; transplantation;  
 KW allergy; systemic lupus erythematosus; haemophilia; Alzheimer's disease;  
 KW graft versus host disease; Lesch-Nyhan syndrome; periodontitis;  
 KW pancreatitis; musculoskeletal disorder; Parkinson's disease;  
 KW Huntington's disease; behavioural disorder; pain; obesity; wound healing;  
 KW neurodegenerative disorder; neuropsychiatric disorder; hypertension;  
 KW growth disorder; reproductive disorder; lung disease.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200216600-A2.  
 PD 28-FEB-2002.  
 XX  
 PD 27-AUG-2001; 2001WO-US26518.  
 XX  
 XX 25-AUG-2000; 2000US-227800P.  
 PR 25-AUG-2000; 2000US-228205P.  
 PR 25-AUG-2000; 2000US-228324P.  
 PR 30-AUG-2000; 2000US-228979P.  
 PR 30-AUG-2000; 2000US-229185P.  
 PR 01-SEP-2000; 2000US-229780P.  
 PR 01-SEP-2000; 2000US-229848P.  
 PR 01-SEP-2000; 2000US-229850P.  
 PR 22-JAN-2001; 2001US-263337P.  
 PR 31-JAN-2001; 2001US-265518P.  
 PR 15-MAR-2001; 2001US-276451P.  
 PR 27-MAR-2001; 2001US-279196P.  
 PR 24-AUG-2001; 2001US-0393398.  
 XX  
 XX (CURA-) CURAGEN CORP.  
 PA Gerlach V, MacDougall JR, Smithson G, Stone DJ, Ellerman K;  
 PI Sytsek KA, Zernhusen BD, Rastelli L, Verney CAM, Paturajan M;  
 PI Tchernev VT, Padigar M, Taupier RJ;  
 XX  
 XX WPI; 2002-292064/33.  
 DR N-PSDB; ABK55573.  
 XX  
 PT New isolated cytoplasmic, nuclear, membrane bound and secreted  
 PT polypeptides, termed NOVX, useful for treating inflammation, autoimmune  
 PT disorders, haemophilia, Lesch-Nyhan syndrome, pancreatitis,  
 PT musculoskeletal disorders -  
 XX  
 PS Claim 1; Page 76-77; 245pp; English.



CC The invention relates to an isolated cytoplasmic, nuclear, membrane bound  
 CC or secreted polypeptide, designated NOVX (actually NOV1, 2a, 2b, 3a, 3b,  
 CC 4, 5a, 5b, 5c, 5d, 5e, 5f, 5g, 5h, 5i, 6, 7 and 8), a variant of NOVX, a  
 CC mature form, or a variant of the mature form of NOVX. Also included  
 CC are a polynucleotide encoding NOVX (or its complement), a vector  
 CC comprising the polynucleotide, a cell comprising the vector, an  
 CC anti-NOVX antibody, determining the presence of NOVX in a sample  
 CC using the antibody, determining the presence of NOVX polynucleotide in  
 CC a sample using a probe which binds to NOVX polynucleotide, identifying a  
 CC agent which binds to NOVX (including modulators of NOVX). NOVX, the  
 CC polynucleotide and the antibody are useful for diagnosing, treating  
 CC or preventing a NOVX-associated disorder selected from cardiomyopathy,  
 CC atherosclerosis, diabetes, a disorder related to cell signal processing  
 CC and metabolic pathway modulation, inflammation, autoimmune disorders,  
 CC scleroderma, transplantation, allergies, systemic lupus erythematosus,  
 CC haemophilia, graft versus host disease, Alzheimer's disease, stroke,  
 CC Lesch-Nyhan syndrome, periodontitis, pancreatitis, musculoskeletal  
 CC disorders, Parkinson's disease, Huntington's disease, behavioural  
 CC disorders, pain, neurodegenerative and neuropsychiatric disorders,  
 CC hypertension, wound healing, obesity, growth and reproductive  
 CC disorders, lung diseases and many other diseases and disorders listed in  
 CC the specification. NOVX, the polynucleotide and the antibody are useful  
 CC in screening assays, detection assays (e.g., chromosomal mapping, tissue  
 CC typing, forensic biology), predictive medicine (e.g., diagnostic assays,  
 CC prognostic assays, monitoring clinical trials and pharmacogenomic), and  
 CC in methods of treatment (e.g., therapeutic and prophylactic). NOVX is  
 CC useful as immunogen to produce antibodies immunospecific for NOVX, as  
 CC vaccines to screen for potential agonist and antagonist compounds, and  
 CC as bait protein in a two-hybrid or three-hybrid assay. The polynucleotide  
 CC is useful in gene therapy, to express NOVX, to detect NOVX mRNA  
 CC or a genetic lesion in a NOVX gene, and to modulate NOVX activity. The  
 CC vector is useful for producing non-human transgenic animals. The antibody  
 CC is useful for isolating, and purifying NOVX and to monitor protein levels  
 CC in tissue as part of a clinical testing procedure. The present sequence  
 CC represents a NOVX protein.

CC Sequence 847 AA;

CC Query Match 89.2%; Score 33; DB 23; Length 847;

CC Best Local Similarity 100.0%; Pred. No. 2.4e+02;

CC Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GETREP 6  
 DB 792 GETREP 797

RESULT 7  
 AAU91288  
 AAU91288 standard; Protein; 857 AA.

AC AAU91288;

DE 18-JUN-2002 (first entry)

XX Human NOV5g protein.  
 XX Human; NOV5; gene therapy; cardiomyopathy; atherosclerosis;  
 XX diabetes; cell signal processing; metabolic pathway modulation;  
 XX inflammation; autoimmune disorder; scleroderma; transplantation;  
 XX allergy; systemic lupus erythematosus; haemophilia; Alzheimer's disease;  
 XX graft versus host disease; Lesch-Nyhan syndrome; periodontitis;  
 XX pancreatitis; musculoskeletal disorder; Parkinson's disease;  
 XX Huntington's disease; behavioural disorder; pain; obesity; wound healing;  
 XX neurodegenerative disorder; neuropsychiatric disorder; hypertension;  
 XX growth disorder; reproductive disorder; lung disease.

OS Homo sapiens.

XX MO200216600-A2.

XX 28-FEB-2002.

XX

PF 27-AUG-2001; 2001WO-US26518.  
 XX 25-AUG-2000; 2000US-227800P.  
 PR 25-AUG-2000; 2000US-228205P.  
 PR 25-AUG-2000; 2000US-228324P.  
 PR 30-AUG-2000; 2000US-228997P.  
 PR 30-AUG-2000; 2000US-229185P.  
 PR 01-SEP-2000; 2000US-229780P.  
 PR 01-SEP-2000; 2000US-229848P.  
 PR 22-JAN-2001; 2001US-263337P.  
 PR 31-JAN-2001; 2001US-265518P.  
 PR 15-MAR-2001; 2001US-276451P.  
 PR 27-MAR-2001; 2001US-279196P.  
 PR 24-AUG-2001; 2001US-0393398.  
 XX (CUBA-) CUBAGEN CORP.  
 PA Gerlach V, MacDougall JR, Smithson G, Stone DJ, Ellerman K;  
 PI Spytek KA, Zernhsen BD, Rastelli L, Verney CAM, Patturajan M;  
 PI Therneve VT, Padigar M, Taupier RJ,  
 XX WPI; 2002-292064/33.  
 DR N-PSDB; ABR55572.  
 PT New isolated cytoplasmic, nuclear, membrane bound and secreted  
 PT polypeptides, termed NOVX, useful for treating inflammation, autoimmune  
 PT disorders, haemophilia, Lesch-Nyhan syndrome, pancreatitis,  
 PT musculoskeletal disorders  
 PS Claim 1; Page 72-73; 245P; English.

XX The invention relates to an isolated cytoplasmic, nuclear, membrane bound  
 CC or secreted polypeptide, designated NOVX (actually NOV1, 2a, 2b, 3a, 3b,  
 CC 4, 5a, 5b, 5c, 5d, 5e, 5f, 5g, 5h, 5i, 6, 7 and 8), a variant of NOVX, a  
 CC mature form, or a variant of the mature form of NOVX. Also included  
 CC are a polynucleotide encoding NOVX (or its complement), a vector  
 CC comprising the polynucleotide, a cell comprising the vector, an  
 CC anti-NOVX antibody, determining the presence of NOVX in a sample  
 CC using the antibody, determining the presence of NOVX polynucleotide in  
 CC a sample using a probe which binds to NOVX polynucleotide, identifying a  
 CC agent which binds to NOVX (including modulators of NOVX). NOVX, the  
 CC polynucleotide and the antibody are useful for diagnosing, treating  
 CC or preventing a NOVX-associated disorder selected from cardiomyopathy,  
 CC atherosclerosis, diabetes, a disorder related to cell signal processing  
 CC and metabolic pathway modulation, inflammation, autoimmune disorders,  
 CC scleroderma, transplantation, allergies, systemic lupus erythematosus,  
 CC haemophilia, graft versus host disease, Alzheimer's disease, stroke,  
 CC Lesch-Nyhan syndrome, periodontitis, pancreatitis, musculoskeletal  
 CC disorders, Parkinson's disease, Huntington's disease, behavioural  
 CC disorders, pain, neurodegenerative and neuropsychiatric disorders,  
 CC hypertension, wound healing, obesity, growth and reproductive  
 CC disorders, lung diseases and many other diseases and disorders listed in  
 CC the specification. NOVX, the polynucleotide and the antibody are useful  
 CC in screening assays, detection assays (e.g., chromosomal mapping, tissue  
 CC typing, forensic biology), predictive medicine (e.g., diagnostic assays,  
 CC prognostic assays, monitoring clinical trials and pharmacogenomic), and  
 CC in methods of treatment (e.g., therapeutic and prophylactic). NOVX is  
 CC useful as immunogen to produce antibodies immunospecific for NOVX, as  
 CC vaccines to screen for potential agonist and antagonist compounds, and  
 CC as bait protein in a two-hybrid or three-hybrid assay. The polynucleotide  
 CC is useful in gene therapy, to express NOVX, to detect NOVX mRNA  
 CC or a genetic lesion in a NOVX gene, and to modulate NOVX activity. The  
 CC vector is useful for producing non-human transgenic animals. The antibody  
 CC is useful for isolating, and purifying NOVX and to monitor protein levels  
 CC in tissue as part of a clinical testing procedure. The present sequence  
 CC represents a NOVX protein.

SO Sequence 857 AA;

Query Match 89.2%; Score 33; DB 23; Length 857;

Best Local Similarity 100.0%; Pred. No. 2.4e+02;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GETREP 6  
 |||||  
 Db 802 GETREP 807

# RESULT 8

AAU91290  
 ID AAU91290 standard; Protein: 905 AA.

AC AAU91290;

DT 18-JUN-2002 (first entry)

DE Human NOV5l protein.

XX Human, NOV5l, gene therapy; cardiomyopathy; atherosclerosis;  
 KW diabetes; cell signal processing; metabolic pathway modulation;  
 KW inflammation; autoimmune disorder; scleroderma; transplantation;  
 KW allergy; systemic lupus erythematosus; haemophilia; Alzheimer's disease;  
 KW graft versus host disease; Leisch-Nyhan syndrome; periodontitis;  
 KW pancreatitis; musculoskeletal disorder; Parkinson's disease;  
 KW Huntington's disease; behavioural disorder; pain; obesity; wound healing;  
 KW neurodegenerative disorder; neuropsychiatric disorder; hypertension;  
 KW growth disorder; reproductive disorder; lung disease.

XX Homo sapiens.

XX NO200216600-A2.

XX 28-FEB-2002.

PF 27-AUG-2001; 2001WO-US26518.

PR 25-AUG-2000; 2000US-227800P.

PR 25-AUG-2000; 2000US-228205P.

PR 25-AUG-2000; 2000US-228324P.

PR 30-AUG-2000; 2000US-228997P.

PR 30-AUG-2000; 2000US-229185P.

PR 01-SEP-2000; 2000US-229780P.

PR 01-SEP-2000; 2000US-229848P.

PR 01-SEP-2000; 2000US-229850P.

PR 22-JAN-2001; 2001US-26337P.

PR 31-JAN-2001; 2001US-265518P.

PR 15-MAR-2001; 2001US-276451P.

PR 27-MAR-2001; 2001US-279196P.

PR 24-AUG-2001; 2001US-0393398.

XX (CURA-) CURAGEN CORP.

XX Gerlach V, MacDougall JR, Smithson G, Stone DJ, Ellerman K;

PI Szytek KA, Zetseren BD, Raselli L, Verney CAM, Patuvarajan M;

PI Tchervet VT, Padigaru M, Taupier RJ;

XX WPI; 2002-292064/33.

DR N-PSDB; ABK55574.

XX New isolated cytoplasmic, nuclear, membrane bound and secreted

PT polypeptides, termed NOVX, useful for treating inflammation, autoimmune

PT disorders, haemophilia, Leisch-Nyhan syndrome, pancreatitis,

PT musculoskeletal disorders

PS Claim 1; Page 81; 245pp; English.

CC an agent which binds to NOVX (including modulators of NOVX). NOVX, the  
 CC polynucleotide and the antibody are useful for diagnosing, treating  
 CC or preventing a NOVX-associated disorder selected from cardiomyopathy,  
 CC atherosclerosis, diabetes, a disorder related to cell signal processing  
 CC and metabolic pathway modulation, inflammation, autoimmune disorders,  
 CC scleroderma, transplantation, allergies, systemic lupus erythematosus,  
 CC haemophilia, graft versus host disease, Alzheimer's disease, stroke,  
 CC Leisch-Nyhan syndrome, periodontitis, pancreatitis, musculoskeletal  
 CC disorders, Parkinson's disease, Huntington's disease, behavioural  
 CC disorders, pain, neurodegenerative and neuropsychiatric disorders,  
 CC hypertension, wound healing, obesity, growth and reproductive  
 CC disorders, lung diseases and many other diseases and disorders listed in  
 CC the specification. NOVX, the polynucleotide and the antibody are useful  
 CC in screening assays, detection assays (e.g., chromosomal mapping, tissue  
 CC typing, forensic biology), predictive medicine (e.g., diagnostic assays,  
 CC prognostic assays, monitoring clinical trials and pharmacogenomic), and  
 CC in methods of treatment (e.g., therapeutic and prophylactic). NOVX is  
 CC useful as immunogen to produce antibodies immunospecific for NOVX, as  
 CC vaccines to screen for potential agonist and antagonist compounds, and  
 CC as bait protein in a two-hybrid or three-hybrid assay. The polynucleotide  
 CC is useful in gene therapy, to express NOVX, to detect NOVX mRNA  
 CC or a genetic lesion in a NOVX gene, and to modulate NOVX activity. The  
 CC vector is useful for producing non-human transgenic animals. The antibody  
 CC is useful for isolating, and purifying NOVX and to monitor protein levels  
 CC in tissue as part of a clinical testing procedure. The present sequence  
 CC represents a NOVX protein.

SO Sequence 905 AA;

Query Match 89.2%; Score 33; DB 23; Length 905;

Best Local Similarity 100.0%; Pred. No. 2.6e+02; Mismatches 0; Gaps 0;

Matches 6; Conservative 0; Indels 0;

OY 1 GETREP 6  
 |||||  
 Db 850 GETREP 855

# RESULT 9

AAU91285  
 ID AAU91285 standard; Protein: 1012 AA.

AC AAU91285;

DT 18-JUN-2002 (first entry)

DE Human NOV5d protein.

XX Human, NOV5d, gene therapy; cardiomyopathy; atherosclerosis;  
 KW diabetes; cell signal processing; metabolic pathway modulation;  
 KW inflammation; autoimmune disorder; scleroderma; transplantation;  
 KW allergy; systemic lupus erythematosus; haemophilia; Alzheimer's disease;  
 KW graft versus host disease; Leisch-Nyhan syndrome; periodontitis;  
 KW pancreatitis; musculoskeletal disorder; Parkinson's disease;  
 KW Huntington's disease; behavioural disorder; pain; obesity; wound healing;  
 KW neurodegenerative disorder; neuropsychiatric disorder; hypertension;  
 KW growth disorder; reproductive disorder; lung disease.

XX Homo sapiens.

XX NO200216600-A2.

XX 28-FEB-2002.

PF 27-AUG-2001; 2001WO-US26518.

PR 25-AUG-2000; 2000US-227800P.

PR 25-AUG-2000; 2000US-228205P.

PR 25-AUG-2000; 2000US-228324P.

PR 30-AUG-2000; 2000US-228997P.

PR 30-AUG-2000; 2000US-229185P.

PR 01-SEP-2000; 2000US-229780P.

PR 01-SEP-2000; 2000US-229848P.

01-SEP-2000; 2000US-229850P.  
PR 22-JAN-2001; 2001US-263337P.  
PR 31-JAN-2001; 2001US-265518P.  
PR 15-MAR-2001; 2001US-276451P.  
PR 27-MAR-2001; 2001US-279196P.  
PR 24-AUG-2001; 2001US-0393398.  
XX  
PA (CURA-) CURAEN CORP.  
PI Gerlach V, MacDougall JR, Smithson G, Stone DJ, Ellerman K,  
PI Sytkew KA, Zernhusen BD, Rastelli L, Verney CAM, Paturajan M;  
PI Tchernev VT, Padiguru M, Taupier RJ,  
DR WPI: 2002-292064/33.  
DR N-PSDB: ABRK55569.  
XX  
PT New isolated cytoplasmic, nuclear, membrane bound and secreted  
PT polypeptides, termed NOVX, useful for treating inflammation, autoimmune  
PT disorders, hemophilia, Leech-Nyhan syndrome, pancreatitis,  
PT musculoskeletal disorders -  
XX  
PS Claim 1; Page 63; 245pp; English.

The invention relates to an isolated cytoplasmic, nuclear, membrane bound or secreted polypeptide, designated NOVX (actually NOV1, 2a, 2b, 3a, 3b, 4, 5a, 5b, 5c, 5d, 5e, 5f, 5g, 5h, 5i, 6, 7 and 8), a variant of NOVX, a mature form, or a variant of the mature form of NOVX. Also included are a polynucleotide encoding NOVX (or its complement), a vector comprising the polynucleotide, a cell comprising the vector, an anti-NOVX antibody, determining the presence of NOVX in a sample using the antibody, determining the presence of NOVX polynucleotide in a sample using a probe which binds to NOVX polynucleotide, identifying a agent which binds to NOVX (including modulators of NOVX). NOVX, the polynucleotide and the antibody are useful for diagnosing, treating or preventing a NOVX-associated disorder selected from cardiomyopathy, atherosclerosis, diabetes, a disorder related to cell signal processing and metabolic pathway modulation, inflammation, autoimmune disorders, scleroderma, transplantation, allergies, systemic lupus erythematosus, hemophilia, graft versus host disease, Alzheimer's disease, stroke, Leech-Nyhan syndrome, peridontitis, pancreatitis, musculoskeletal disorders, Parkinson's disease, Huntington's disease, behavioural disorders, pain, neurodegenerative and neuropsychiatric disorders, hypertension, wound healing, obesity, growth and reproductive disorders, lung diseases and many other diseases and disorders listed in the specification. NOVX, the polynucleotide and the antibody are useful in screening assays, detection assays (e.g., chromosomal mapping, tissue typing, forensic biology), predictive medicine (e.g., diagnostic assays, prognostic assays, monitoring clinical trials and pharmacogenomic), and in methods of treatment (e.g., therapeutic and prophylactic). NOVX is useful as immunogen to produce antibodies immunospecific for NOVX, as vaccines to screen for potential agonist and antagonist compounds, and as bait protein in a two-hybrid or three-hybrid assay. The polynucleotide is useful in gene therapy, to express NOVX, to detect NOVX mRNA or a genetic lesion in a NOVX gene, and to modulate NOVX activity. The vector is useful for producing non-human transgenic animals. The antibody is useful for isolating, and purifying NOVX and to monitor protein levels in tissue as part of a clinical testing procedure. The present sequence represents a NOVX protein.

Sequence 1012 AA;

Query Match 89.2%; Score 33; DB 23; Length 1012;  
Best Local Similarity 100.0%; Pred. No. 2.9e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0

OY 1 GETREP 6  
| | | | |  
Db 957 GETREP 962

RESULT 10  
AAI36828  
ID AAI36828 standard; Protein; 102 AA.

XX	AAV36828;
AC	
XX	07-OCT-1999 (first entry)
DT	
XX	
DE	Chlamydia trachomatis lipoprotein sequence.
XX	
KW	Vaccine; eye disease; conventional trachoma; nonendemic trachoma; paratrachoma; inclusion conjunctivitis; genital disease; perithenaritis; nongonococcal urethritis; epididymitis; cervicitis; salpingitis; batholinits; pneumopathy; venereal lymphogranulomatosis.
KX	
XX	Chlamydia trachomatis.
OS	
XX	
FN	M09928475-A2.
XX	
PD	10-JUN-1999.
XX	
PF	27-NOV-1998; 98MO-IB01939.
XX	
PR	04-NOV-1998; 98US-0107077. 28-NOV-1997; 97FR-0015041. 17-DEC-1997; 97FR-0016034.
XX	
PA	(GEST ) GENSET.
PI	Giffais R;
DR	MP1, 1999-371125/31.
XX	
FT	Genome sequence of Chlamydia trachomatis
PS	
XX	Disclosure; Page 708; 1755DP; English.
CC	AAV36754-Y37949 are encoded by open reading frames (ORFs) of the genome of Chlamydia trachomatis (see AAV36754). The polypeptides can be used as vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences can also be used to control growth of the microorganism. Chlamydia trachomatis is responsible for a large number of diseases, e.g. eye diseases such as conventional trachoma, nongonorrheic trachoma, paratrachoma, and inclusion conjunctivitis; genital diseases such as nongonococcal urethritis, epididymitis, cervicitis, salpingitis, perithenaritis, batholinits; pneumopathy in breast feeding infants; and venereal lymphogranulomatosis. The polypeptides of the invention may be of use in treating these diseases.
CC	
CC	
CC	
SQ	Sequence 102 AA;
Query Match	86.5%; Score 32; DB 20; Length 102;
Best Local Similarity	85.7%; Pred. No. 39;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;	
YY	1 GETREPL 7
DB	82 GENREPL 88
RESULT 11	
AAU43585	
ID	AAU43585 standard; Protein; 114 AA.
XX	
AAU43585;	
XX	
DT	27-FEB-2002 (first entry)
XX	
DE	Propionibacterium acnes immunogenic protein #4481.
XX	
KW	SAPHO syndrome; synovitis; acne; pustulosis; hyperkeratosis; osteomyelitis; osteitis; endophthalmitis; bone; joint; central nervous system; ELISA; inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay; dermatological; osteopathic; neuroprotectant.
KX	
XX	Propionibacterium acnes.
SQ	

PN MO200101581-A2.  
 XX  
 PD 01-NOV-2001.  
 XX  
 PF 20-APR-2001; 2001WO-US12865.  
 XX  
 PR 21-APR-2000; 2000US-199047P.  
 PR 02-JUN-2000; 2000US-208841P.  
 PR 07-JUL-2000; 2000US-216747P.  
 XX  
 PA (CORI-) CORIXA CORP.  
 PI Skeiky YAM, Pearing DH, Mitcham JU, Wang SS, Bhatia A;  
 PI L'maisonmeuve J, Zhang Y, Jen S, Carter D;  
 DR WPI; 2001-616774/71.  
 DR N-PSDB; NA559521.  
 XX  
 PT Propionibacterium acnes polypeptides and nucleic acids useful for  
 PT vaccinating against and diagnosing infections, especially useful for  
 PT treating acne vulgaris -  
 XX  
 PS Example 1; SEQ ID NO 4780; 1069pp; English.  
 XX  
 CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic  
 CC polypeptides. The proteins and their associated DNA sequences are used in  
 CC the treatment, prevention and diagnosis of medical conditions caused by  
 CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,  
 CC pustulosis, hyperostosis and osteomyelitis), uveitis and endophthalmitis.  
 CC P. acnes is also involved in infections of bone, joints and the central  
 CC nervous system, however it is particularly involved in the inflammatory  
 CC lesions associated with acne vulgaris. A method for detecting the  
 CC presence or absence of P. acnes in a patient comprises contacting a  
 CC sample with a binding agent that binds to the proteins of the invention  
 CC and determining the amount of bound protein in the sample. The  
 CC polypeptides may be used as antigens in the production of antibodies  
 CC specific for P. acnes proteins. These antibodies can be used to  
 CC downregulate expression and activity of P. acnes polypeptides and  
 CC therefore treat P. acnes infections. The antibodies may also be used as  
 CC diagnostic agents for determining P. acnes presence, for example, by  
 CC enzyme linked immunosorbent assay (ELISA).  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 CC  
 CC  
 SQ Sequence 114 AA;  
 Query Match: 86.5%; Score 32; DB 22; Length 114;  
 Best Local Similarity 85.7%; Pred. No. 44;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 OY 1 GETREP.L 7  
 | | | | |  
 Db 55 GRTREP.L 61  
 RESULT 12  
 AAU64543  
 ID AAU64543 standard; Protein; 179 AA.  
 XX  
 XX AAU64543;  
 DT 27-FEB-2002 (first entry)  
 XX  
 DE Propionibacterium acnes immunogenic protein #25439.  
 XX  
 KW SAPHO syndrome; synovitis; acne; pustulosis; hyperostosis; osteomyelitis;  
 KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
 KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
 KW dermatological; osteopathic; neuroprotectant.  
 XX  
 OS Propionibacterium acnes.

```

XX FN WO200181561-A2.
XX FD
XX ED 01-NOV-2001.
XX PF
XX 20-APR-2001; 2001WO-US12865.
XX PR
XX 21-APR-2000; 2000US-199047P.
XX PR 02-JUN-2000; 2000US-208841P.
XX ER 07-JUL-2000; 2000US-216747P.
XX PA
XX (CORI-) CORIXA CORP.
XX PI
XX PI Skeiky YAM, Persing DH, Mitcham JL, Wang SS, Bhatia A;
XX PI L'maisonneuve U, Zhang Y, Jen S, Carter D;
XX DR
XX WPI; 2001-616774/71.
XX DR
XX N-PSDB; AAS59645.
XX PT
XX PT Propionibacterium acnes polypeptides and nucleic acids useful for
XX PT vaccinating against and diagnosing infections, especially useful for
XX PT treating acne vulgaris -
XX PS
XX PS Example 1; SEQ ID No 25738; 10699p; English.
XX CC
XX CC Sequences AAU391905-AAU68017 represent Propionibacterium acnes immunogenic
XX CC polypeptides. The proteins and their associated DNA sequences are used in
XX CC the treatment, prevention and diagnosis of medical conditions caused by
XX CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
XX CC pustulosis, hyperostosis and osteomyelitis), uveitis and endophthalmitis.
XX CC P. acnes is also involved in infections of bone, joints and the central
XX CC nervous system, however it is particularly involved in the inflammatory
XX CC lesions associated with acne vulgaris. A method for detecting the
XX CC presence or absence of P. acnes in a patient comprises contacting a
XX CC sample with a binding agent that binds to the proteins of the invention
XX CC and determining the amount of bound protein in the sample. The
XX CC polypeptides may be used as antigens in the production of antibodies
XX CC specific for P. acnes proteins. These antibodies can be used to
XX CC downregulate expression and activity of P. acnes polypeptides and
XX CC therefore treat P. acnes infections. The antibodies may also be used as
XX CC diagnostic agents for determining P. acnes presence, for example, by
XX CC enzyme linked immunosorbent assay (ELISA).
XX CC Note: The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ
XX SQ Sequence 179 AA;
XX
XX Query March 86.5%; Score 32; DB 22; Length 179;
XX Best Local Similarity 85.7%; Pred. NO. 71;
XX Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX QY 1 GETREPL 7
XX |||||
XX |||||
XX Db 84 GETRSPL 90
XX
XX RESULT 13
XX ID AAB44981
XX AAB44981 standard; Protein; 339 AA.
XX XX
XX AAB44981;
XX XX
XX 12-FEB-2001 (first entry)
XX XX
XX Human secreted protein encoded by gene 10 homologue.
XX DE
XX Secreted protein; human; immunosuppressive; antiarthritic; antirheumatic;
XX RW antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective;
XX RW neurotropic; neuroprotective; antibacterial; virucide; fungicide;
XX RW ophtalmological; gene therapy; treatment; autoimmune disease; infection;
XX RW cardiovascular disorder; nervous system disorder; ocular disorder;
XX RW wound healing; epithelial cell proliferation; skin aging; mental state;

```

KM Transplantation; metabolism modulation.  
 XX Homo sapiens.  
 OS  
 XX MO200055200-A1.  
 PN  
 XX 21-SEP-2000.  
 PD  
 XX 09-MAR-2000; 2000MO-US06042.  
 XX  
 PF  
 XX 12-MAR-1999; 99US-0124143.  
 PR 03-DEC-1999; 99US-0168663.  
 XX  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 PA  
 PI Rosen CA, Ruben SM, Komatsuolis G;  
 XX WPI; 2000-656008/63.  
 DR  
 XX  
 XX Isolated human secretory proteins, nucleic acids encoding them and  
 PT antibodies directed against them, useful for diagnosing and treating  
 PT disorders related to the proteins such as cancer, Alzheimer's disease  
 PT and Parkinsons -  
 XX  
 PS Disclosure; Page 407-408; 453pp; English.  
 XX  
 XX This invention describes a novel isolated polypeptide (I) and its  
 CC encoding nucleic acid molecule (II) which have immunosuppressive,  
 CC antitumor, antirheumatic, antiproliferative, cytostatic, cardiant,  
 CC vasotropic, cerebroprotective, neurotropic, neuroprotective, antibacterial,  
 CC virucide, fungicide and ophthalmological activity and which can be used  
 CC for gene therapy. (I) and (II) are used to prevent, treat or ameliorate  
 CC a medical condition in e.g. humans, mice, rabbits, goats, horses, cats,  
 CC dogs, chickens or sheep. (I) and (II) are also used in diagnosing a  
 CC pathological condition or susceptibility to a pathological condition. The  
 CC antibodies to (II) can also be used in alleviating symptoms associated  
 CC with the disorders and in diagnostic immunoassays e.g. radioimmunoassays  
 CC or enzyme linked immunosorbent assays (ELISA). Disorders which are  
 CC diagnosed or treated include autoimmune diseases e.g. rheumatoid  
 CC arthritis, hyperproliferative disorders e.g. neoplasms of the breast or  
 CC liver, cardiovascular disorders e.g. cardiac arrest, cerebrovascular  
 CC disorders e.g. cerebral ischemia, angiodenesis, nervous system disorders  
 CC e.g. Alzheimer's disease, infections caused by bacteria, viruses and  
 CC fungi and ocular diseases e.g. corneal infection. The polypeptides can  
 CC also be used to aid wound healing and epithelial cell proliferation, to  
 CC prevent skin aging due to sunburn, to maintain organs before  
 CC transplantation, for supporting cell culture of primary tissues, to  
 CC regenerate tissues and in chemotaxis. The polypeptides are used to  
 CC modulate mammalian metabolism, to change mammal's mental state or  
 CC physical state by influencing biohythms circadian rhythms, depression  
 CC tendency for violence tolerance for pain, reproductive capabilities,  
 CC hormonal or endocrine levels, appetite, libido, memory, stress or other  
 CC cognitive qualities, as a food additive or preservative, such as to  
 CC increase or decrease storage capabilities, fat content, lipid, protein,  
 CC carbohydrates, vitamins, minerals, cofactors or other nutritional  
 CC components.  
 CC  
 XX Sequence 339 AA;  
 SQ  
 Query Match 86.5%; Score 32; DB 21; Length 339;  
 Best Local Similarity 85.7%; Pred. No. 1.4e+02;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 GENTREPL 7  
 DB 19 GNTREPL 25  
 RESULT 14  
 AAB95588  
 ID AAB95588 standard; Protein; 611 AA.  
 XX  
 AC AAB95588;

XX  
 DT 26-JUN-2001 (first entry)  
 XX  
 DE Human protein sequence SEQ ID NO:18257.  
 XX  
 XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.  
 KM  
 XX Homo sapiens.  
 OS  
 XX  
 XX EP1074617-A2.  
 PN  
 XX 07-FEB-2001.  
 PD  
 XX 28-JUL-2000; 2000EP-0116126.  
 XX  
 PF  
 XX 29-JUL-1999; 99JP-0248036.  
 PR 27-AUG-1999; 99JP-0300253.  
 PR 11-JAN-2000; 2000JP-0118776.  
 PR 02-MAY-2000; 2000JP-0183767.  
 PR 09-JUN-2000; 2000JP-0241899.  
 XX  
 XX (HELI-) HELIX RES INST.  
 PA  
 XX Oca T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
 XX WPI; 2001-318749/34.  
 DR  
 XX  
 XX Primer sets for synthesizing polynucleotides, particularly the 5602  
 PT full-length cDNAs defined in the specification, and for the detection  
 PT and/or diagnosis of the abnormality of the proteins encoded by the  
 PT full-length cDNAs -  
 XX  
 PS Claim 8; SEQ ID 18257; 2537pp + CD ROM; English.  
 XX  
 XX The present invention describes primer sets for synthesizing 5602  
 CC full-length cDNAs defined in the specification. Where a primer set  
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
 CC to the complementary strand of a polynucleotide which comprises one of  
 CC the 5602 nucleotide sequences defined in the specification, where the  
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 CC of an oligonucleotide comprising a sequence complementary to the  
 CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises a 3'-end sequence, where the  
 CC oligonucleotide comprises at least 15 nucleotides and the combination of  
 CC the 5'-end sequence/3'-end sequence is selected from those defined in  
 CC the specification. The primer sets can be used in antisense therapy and  
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
 CC represent oligonucleotides, all of which are used in the exemplification  
 CC of the present invention.  
 CC  
 XX Sequence 611 AA;  
 SQ  
 Query Match 86.5%; Score 32; DB 22; Length 611;  
 Best Local Similarity 85.7%; Pred. No. 2.7e+02;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 GENTREPL 7  
 DB 291 GNTREPL 297  
 RESULT 15  
 ABG25333  
 ID ABG25333 standard; Protein; 846 AA.  
 XX

```

AC  ABG25333;
XX
XX  18-FEB-2002 (first entry)
DT
XX
XX  Novel human diagnostic protein #25324.
DE
XX
XX  Human; chromosome mapping; gene mapping; gene therapy; forensic;
KM  food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX  Homo sapiens.
OS
XX  WO200175067-A2.
XX  11-OCT-2001.
XX
XX  30-MAR-2001; 2001WO-US06631.
XX
XX  31-MAR-2000; 2000US-0540217.
XX  23-AUG-2000; 2000US-0649167.
XX
XX  (HYSE-) HYSEQ INC.
XX
XX  Dmanac RT, Liu C, Tang YT;
XX
XX  WPI; 2001-639362/73.
XX  N-PSDB; AAS89520.
XX
XX  New isolated polynucleotide and encoded polypeptides, useful in
PT  diagnostics, forensics, gene mapping, identification of mutations
PT  responsible for genetic disorders or other traits and to assess
PT  biodiversity -
XX
XX  Claim 20; SEQ ID No 55692; 103bp; English.
XX
XX  The invention relates to isolated polynucleotide (I) and
XX  polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX  polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX  and gene mapping, and in recombinant production of (II). The
XX  polynucleotides are also used in diagnostics as expressed sequence tags
XX  for identifying expressed genes. (I) is useful in gene therapy techniques
XX  to restore normal activity of (II) or to treat disease states involving
XX  (II). (II) is useful for generating antibodies against it, detecting or
XX  quantitating a polypeptide in tissue, as molecular weight markers and as
XX  a food supplement. (II) and its binding partners are useful in medical
XX  imaging of sites expressing (II). (I) and (II) are useful for treating
XX  disorders involving aberrant protein expression or biological activity.
XX  The polypeptide and polynucleotide sequences have applications in
XX  diagnostics, forensics, gene mapping, identification of mutations
XX  responsible for genetic disorders or other traits to assess biodiversity
XX  and to produce other types of data and products dependent on DNA and
XX  amino acid sequences. ABG00010-ABG30377 represent novel human
XX  diagnostic amino acid sequences of the invention.
XX  Note: The sequence data for this patent did not appear in the printed
XX  specification, but was obtained in electronic format directly from WIPO
XX  at ftp.wipo.int/pub/published_pct_sequences.
XX
XX  Sequence 846 AA;
SQ

```

Query Match 86.5%; Score 32; DB 22; Length 846;  
Best Local Similarity 85.7%; Pred. No. 3.9e+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

QY  1 GETREPL 7
   | | | | |
Db  746 GNTREPL 752

```

Search completed: January 29, 2003, 13:59:51  
Job time : 26.3077 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2003, 13:57:01 ; Search time 9.87179 Seconds  
(without alignments)  
68.168 Million cell updates/sec

Title: US-09-807-949a-108

Perfect score: 37

Sequence: 1 GETREPL 7

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database:

1: PIR 73: \*  
2: p1r1: \*  
3: p1r2: \*  
4: p1r3: \*  
5: p1r4: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	33	89.2	613	2	G75085 indolepyruvate fer
2	33	89.2	613	2	G75085 probable indolepyr
3	33	89.2	1121	2	D71055 hypothetical prote
4	32	86.5	473	2	G84312 glycine dehydrogen
5	32	86.5	736	2	H75460 conserved hypothet
6	32	86.5	781	1	A49794 DNA topoisomerase
7	32	86.5	796	2	A81997 DNA topoisomerase
8	32	86.5	796	2	D81225 DNA gyrase chain B
9	31	83.8	144	2	B84219 hypothetical prote
10	31	83.8	168	2	T27087 hypothetical prote
11	31	83.8	274	2	S60972 hypothetical prote
12	31	83.8	300	2	AB1188 acetyltransferase
13	31	83.8	325	2	T36485 hypothetical prote
14	31	83.8	390	2	H70508 probable mmp prote
15	31	83.8	405	2	A10006 DNA/pantothenate m
16	31	83.8	407	2	AC0971 conserved hypothet
17	31	83.8	430	1	A65165 pantothenate metab
18	31	83.8	430	2	B91193 pantothenate metab
19	31	83.8	430	2	C86040 pantothenate metab
20	31	83.8	447	2	T04702 pantothenate metab
21	31	83.8	451	2	T35390 probable aminocan
22	30	81.1	88	2	T47139 hypothetical prote
23	30	81.1	130	2	G86804 cytidine deaminase
24	30	81.1	133	2	S77489 ribosomal protein
25	30	81.1	135	2	F72494 hypothetical prote
26	30	81.1	202	2	A69289 hypothetical prote
27	30	81.1	274	2	D75362 Far-related protei
28	30	81.1	277	2	I47162 Ig gamma 4 chain c
29	30	81.1	328	2	I47160 Ig gamma 2b chain

30	30	81.1	328	2	I47159 Ig gamma 2a chain
31	30	81.1	342	2	AF3576 iron(III) diclrat
32	30	81.1	414	2	JC4921 double C2 protein
33	30	81.1	489	2	H97888 glycerol kinase 2
34	30	81.1	489	2	AE2810 glycerol kinase 2
35	30	81.1	502	1	S49205 betaine-aldehyde d
36	30	81.1	611	2	AH2957 hypothetical ABC t
37	30	81.1	611	2	F98325 hypothetical ABC t
38	30	81.1	678	2	G75524 translation elonga
39	30	81.1	862	2	T07775 lipoxigenase (EC 1
40	30	81.1	3588	2	I40485 surfactin syntheta
41	29	78.4	155	2	B83771 hypothetical prote
42	29	78.4	188	2	T46272 hypothetical prote
43	29	78.4	208	2	T26689 hypothetical prote
44	29	78.4	273	2	E70939 hypothetical prote
45	29	78.4	277	2	D82005 50S ribosomal prot

#### ALIGNMENTS

##### RESULT 1

G75085

indolepyruvate ferredoxin oxidoreductase, chain alpha (lora-2) PAB0718 - Pyrococcus abyssi

C:Species: Pyrococcus abyssi

C>Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Jun-2000

C:Accession: G75085

R:anonymous, Genoscope

Submitted to the EMBL Data Library, July 1999

A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome struc

A:Reference number: A75001

A:Accession: G75085

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-613 <RAM>

A:Experimental source: strain Oresy

C:Genetics:

A:Gene: lora-2; PAB0718

C:Superfamily: indolepyruvate synthase alpha chain; ferredoxin 2[4Fe-4S] homology

F;553-609/Domain: ferredoxin 2[4Fe-4S] homology <FER>

Query Match 89.2%; Score 33; DB 2; Length 613;

Best local similarity 100.0%; Pred. No. 29; Mismatches 0; Indels 0; Gaps 0;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GETREP 6

Db 547 GETREP 552

##### RESULT 2

D71055 probable indolepyruvate ferredoxin oxidoreductase alpha subunit - Pyrococcus horikoshii

C:Species: Pyrococcus horikoshii

C>Date: 14-Aug-1998 #sequence\_revision 14-Aug-1998 #text\_change 17-Mar-2000

C:Accession: D71055

R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Hainawa, Y.; Hino, Y.; Yamamoto, S.; Sekine

M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi

DNA Res. 5, 55-76, 1998

A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic ar

A:Reference number: A71000; MUID:98344137; PMID:9679194

A:Accession: D71055

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-618 <RAM>

A:Cross-reference: GB:AP000005; NID:g3236132; PIDN:BA30238.1; PID:dl031181; PID:g32575'

A:Experimental source: strain OT3

A>Note: this accession replaces an interim accession for a sequence replaced by GenBank

C:Genetics:

A:Gene: PH138

C:Superfamily: unassigned ferredoxin 2[4Fe-4S]-related proteins; ferredoxin 2[4Fe-4S] non

F;558-614/Domain: ferredoxin 2[4Fe-4S] homology <FER4>

Query Match 89.2%; Score 33; DB 2; Length 618;  
C/Accession: H75460  
Best Local Similarity 100.0%; Pred. No. 29;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GETREP 6  
Db 552 GETREP 557

## RESULT 3

T47970  
hypothetical protein F15G16.170 - Arabidopsis thaliana

C/Species: Arabidopsis thaliana (mouse-ear cress)

C/Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 20-Apr-2000

C/Accession: T47970

R/De Haan, M.; Maarse, A.C.; Grivell, L.A.; Mewes, H.W.; Lemcke, K.; Mayer, K.F.X.; Quer  
submitted to the Protein Sequence Database, January 2000

A/Reference number: 224480

A/Accession: T47970

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-1121 <DEH>

A/Cross-references: EMBL:AL132959

A/Experimental source: cultivar Columbia; BAC clone F15G16

C/Genetics:

A/Map position: 3

A/Intons: 754/1; 960/3; 1031/3

A/Note: F15G16.170

Query Match 89.2%; Score 33; DB 2; Length 1121;  
Best Local Similarity 100.0%; Pred. No. 57;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GETREP 6  
Db 660 GETREP 665

## RESULT 4

G84312

glycine dehydrogenase subunit 2 [imported] - Halobacterium sp. NRC-1

C/Species: Halobacterium sp. NRC-1

C/Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001

C/Accession: G84312

R/Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Bergquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.  
; Leitchauer, B.; Keller, K.; Cruz, R.; Danon, M.U.; Hough, D.W.; Maddocks, D.G.; Jaldic  
Jung, K.H.; Alam, M.; Freitas, T.  
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000

A/Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Li

A/Title: Genome sequence of Halobacterium species NRC-1.

A/Reference number: A84160; MUID:20504483; PMID:11016950

A/Accession: G84312

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-473 <STO>

A/Cross-references: GB:AE004437; NID:G10581077; PID:AA019867.1; GSPDB:GN00138

C/Genetics:

A/Gene: gcvp2

Query Match 86.5%; Score 32; DB 2; Length 473;  
Best Local Similarity 85.7%; Pred. No. 37;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GETREP 7  
Db 14 GETREP 20

## RESULT 5

H75460

conserved hypothetical protein - Deinococcus radiodurans (strain R1)

C/Species: Deinococcus radiodurans

C/Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 18-Aug-2000

C/Accession: H75460

R/White, O.; Bisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.; B  
; M.; Shen, M.; Vamathavan, J.U.; Lam, P.; McDonald, L.; Uterback, T.; Zalewski, C.; Mei  
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
Science 286, 1571-1577, 1999

A/Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.

A/Reference number: A75250; MUID:20036896; PMID:10567266

A/Accession: H75460

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-736 <WHI>

A/Cross-references: GB:AE001943; GB:AE000513; NID:96458624; PID:AA010482.1; PID:96458624

A/Experimental source: strain R1

C/Genetics:

A/Map position: 1

C/Superfamily: Haloflex hypothetical protein 4 (gyrB region)

Query Match 86.5%; Score 32; DB 2; Length 736;  
Best Local Similarity 85.7%; Pred. No. 60;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GETREP 7  
Db 605 GETREP 611

## RESULT 6

A49794

DNA topoisomerase (ATP-hydrolyzing) (EC 5.99.1.3) chain B, nalidixic acid-resistant - Nei

C/Species: Neisseria gonorrhoeae

C/Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999

C/Accession: A49794

R/Stein, D.C.; Danaher, R.U.; Cook, T.M.

Antimicrob. Agents Chemother. 35, 622-626, 1991

A/Title: Characterization of a gyrB mutation responsible for low-level nalidixic acid res

A/Reference number: A49794; MUID:91298684; PMID:1906260

A/Accession: A49794

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-781 <STE>

A/Cross-references: GB:M59981

C/Superfamily: DNA topoisomerase (ATP-hydrolyzing) chain B

C/Keywords: isomerase

Query Match 86.5%; Score 32; DB 1; Length 781;  
Best Local Similarity 85.7%; Pred. No. 64;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GETREP 7  
Db 152 GETREP 158

## RESULT 7

A81997

DNA topoisomerase (ATP-hydrolyzing) (EC 5.99.1.3) B NMA0056 [imported] - Neisseria mening

C/Species: Neisseria meningitidis

C/Date: 05-May-2000 #sequence\_revision 05-May-2000 #text\_change 02-Feb-2001

C/Accession: A81997

R/Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel

; Holtroyd, S.; Jorgels, K.; Leather, S.; Mout, S.; Mungall, K.; Quail, M.A.; Rajandream,

Nature 404, 502-506, 2000

A/Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.

A/Reference number: A81775; MUID:20222556; PMID:10761919

A/Accession: A81997

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-796 <PAR>

A/Cross-references: GB:AL162752; GB:AL157959; NID:97378778; PID:CA083373.1; PID:9737883

A/Experimental source: serogroup A, strain Z2491

C/Genetics:



A:Gene: GYTB; NMA0056  
 C:Superfamily: DNA topoisomerase (ATP-hydrolyzing) chain B  
 C:Keywords: isomerase

Query Match 86.5%; Score 32; DB 2; Length 796;  
 Best Local Similarity 85.7%; Pred. No. 65;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GETREPL 7  
 |||||  
 Db 152 GETREPL 158

## RESULT 8

DNA gyrase chain B NMB0212 [imported] - Neisseria meningitidis (strain MC58 serogroup B)

C:Species: Neisseria meningitidis  
 C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 19-Jan-2001

C:Accession: D81225

R:Retelid, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.  
 Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;  
 Ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Maignani, V.; Pizzi, M.

A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Noxon, E.R.; Rappuoli, R.; Ve  
 A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.  
 A:Reference number: A81000; PMID:20175755; PMID:10710307

A:Accession: D81225

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-796 <TEXT>

A:Cross-references: GB:AE002378; GB:AE002098; NID:g1225428; PIDN:AAF40668.1; PID:g722543

A:Experimental source: serogroup B, strain MC58

C:Genetics:

A:Gene: NMB0212

C:Superfamily: DNA topoisomerase (ATP-hydrolyzing) chain B

Query Match 86.5%; Score 32; DB 2; Length 796;  
 Best Local Similarity 85.7%; Pred. No. 65;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GETREPL 7  
 |||||  
 Db 152 GETREPL 158

## RESULT 9

hypothetical protein Vng0613h [imported] - Halobacterium sp. NRC-1

C:Species: Halobacterium sp. NRC-1

C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001

C:Accession: E84219

R:Ng, W.V.; Kennedy, S.P.; Mahairs, G.G.; Bergquist, B.; Pan, M.; Shukla, H.D.; Laaky, S.  
 ; Leitauer, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jabic  
 Jung, K.H.; Alam, M.; Freitae, T.

A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ehardt, H.; Lowe, T.M.; Li  
 A:Title: Genome sequence of Halobacterium species NRC-1.

A:Reference number: A84160; PMID:20504483; PMID:11016550

A:Accession: E84219

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-144 <STO>

A:Cross-references: GB:AE004437; NID:g10580207; PIDN:AA619121.1; GSPDB:GN00138

C:Genetics:

A:Gene: VNG0613H

Query Match 83.8%; Score 31; DB 2; Length 144;  
 Best Local Similarity 71.4%; Pred. No. 16;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 GETREPL 7  
 |||||  
 Db 134 GQTRDPL 140

## RESULT 10

T27087

hypothetical protein Y51B9A.8 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C:Accession: T27087

R:Barlow, K.

A:Submitted to the EMBL Data Library, June 1998

A:Reference number: Z20308

A:Accession: T27087

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-168 <WIL>

A:Cross-references: EMBL:AL023845; PIDN:CAA19536.1; GSPDB:GN00020; CESP:Y51B9A.8

A:Experimental source: clone Y51B9A

C:Genetics:

A:Gene: CESP:Y51B9A.8

A:Map position: 2

A:Introns: 30/1; 68/1; 136/2

QY 2 ETRERPL 7  
 |||||  
 Db 156 ETRERPL 161

## RESULT 11

hypothetical protein YNL155w - yeast (Saccharomyces cerevisiae)

N:Alternate names: hypothetical protein N1751

C:Species: Saccharomyces cerevisiae

C:Date: 15-Feb-1996 #sequence\_revision 01-Mar-1996 #text\_change 19-Apr-2002

C:Accession: S60972; S63107; S63819

R:Nasr, F.; Becam, A.M.; Herbert, C.J.

A:Submitted to the EMBL Data Library, October 1995

A:Description: The sequence of 36.8 kb from the left arm of chromosome XIV reveals 24 com

tonic dystrophy kinase.

A:Reference number: S60972

A:Accession: S60972

A:Molecule type: DNA

A:Residues: 1-274 <NAS>

A:Cross-references: EMBL:X92517; NID:g1050783; PID:g1050797

R:Nasr, F.; Becam, A.M.; Herbert, C.

A:Submitted to the Protein Sequence Database, April 1996

A:Reference number: S62967

A:Accession: S63107

A:Molecule type: DNA

A:Residues: 1-274 <NAF>

A:Cross-references: EMBL:X92517; NID:g1050783; PIDN:CAA63284.1; PID:g1050797

A>Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1995

C:Genetics:

A:Cross-references: SGD:S0005099

A:Map position: 14L

A>Note: YNL155w

Query Match 83.8%; Score 31; DB 2; Length 274;  
 Best Local Similarity 100.0%; Pred. No. 33;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 ETRREPL 7  
 Db 98 ETRREPL 103

## RESULT 12

AB3188  
 acetyltransferase [imported] - Agrobacterium tumefaciens (strain C58, Dupont) plasmid AT  
 C:Species: Agrobacterium tumefaciens  
 C:Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 11-Jan-2002  
 C:Accession: AB3188  
 R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Moo, I.  
 erage, G.; Gillet, M.; Grant, C.; Guenther, D.; Kuyavin, T.; Levy, R.; Li, M.; McClell  
 ; Karp, P.; Romero, P.; Zhang, S.  
 Science 294, 2317-2323, 2001  
 A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,  
 ster, E.W.  
 A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
 A:Reference number: AB2577; PMID:11743193  
 A:Accession: AB3188  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-300 <KUR>  
 A:Cross-references: GB:AE008687; PIDN:AAL45920.1; PID:G17743667; GSPDB:GN00188  
 C:Genetics:  
 A:Experimental source: strain C58 (Dupont)  
 A:Gene: Atus5231  
 A:Genome: plasmid

Query March 83.8%; Score 31; DB 2; Length 300;  
 Best Local Similarity 85.7%; Pred. No. 37;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 GETREPL 7  
 Db 284 GETREPL 290

## RESULT 13

T36495  
 hypochelical protein SCGD3.15 - Streptomyces coelicolor  
 C:Species: Streptomyces coelicolor  
 C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 04-Mar-2000  
 C:Accession: J36495  
 R:Saunders, D.C.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.  
 submitted to the EMBL Data Library, July 1999  
 A:Reference number: Z21608  
 A:Accession: J36495  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-325 <SAU>  
 A:Cross-references: EMBL:AL096822; PIDN:GAB46937.1; GSPDB:GN00070; SCOEDB:SCGD3.15  
 C:Genetics:  
 A:Experimental source: strain A3(2)  
 A:Gene: SCOEDB:SCGD3.15  
 C:Superfamily: Streptomyces coelicolor hypothetical protein SCGD3.15

Query Match 83.8%; Score 31; DB 2; Length 325;  
 Best Local Similarity 85.7%; Pred. No. 40;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 GETREPL 7  
 Db 274 GETREPL 280

## RESULT 14

H70508  
 probable mrp protein - Mycobacterium tuberculosis (strain H37RV)  
 C:Species: Mycobacterium tuberculosis  
 C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000

C:Accession: H70508  
 R:Cole, S.T.; Broesch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.  
 ; Connor, R.; Davies, R.; Devlin, K.; Feltham, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;  
 Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
 Nature 393, 537-544, 1998  
 A:Authors: Sgares, R.; Suleston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
 A:Reference number: A70500; MUID:98295987; PMID:9634230  
 A:Accession: H70508  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-390 <COL>  
 A:Cross-references: GB:Z98260; GB:AL123456; NID:g3261826; PIDN:GAB10925.1; PID:g3261828  
 A:Experimental source: strain H37RV  
 C:Genetics:  
 A:Gene: mrp  
 C:Superfamily: conserved probable membrane protein YII003w

Query Match 83.8%; Score 31; DB 2; Length 390;  
 Best Local Similarity 71.4%; Pred. No. 50;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GETREPL 7  
 Db 110 GETREPL 116

## RESULT 15

A10006  
 DNA/pantothenate metabolism flavoprotein [imported] - Yersinia pestis (strain C092)  
 C:Species: Yersinia pestis  
 C:Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 09-Nov-2001  
 C:Accession: A10006  
 R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.;  
 deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;  
 11, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell, B.  
 Nature 413, 523-527, 2001  
 A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.  
 A:Reference number: AB0001; MUID:21470413; PMID:11586360  
 A:Accession: A10006  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-405 <KUR>  
 A:Cross-references: GB:AL590842; PIDN:CAC88915.1; PID:G15978163; GSPDB:GN00175  
 C:Genetics:  
 A:Gene: dfp  
 C:Superfamily: pantothenate metabolism flavoprotein dfp

Query Match 83.8%; Score 31; DB 2; Length 405;  
 Best Local Similarity 85.7%; Pred. No. 52;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 GETREPL 7  
 Db 198 GETREPL 204

Search completed: January 29, 2003, 14:03:24  
 Job time : 11.8718 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2003, 13:51:40 ; Search time 4.04615 Seconds  
(without alignments)  
59.910 Million cell updates/sec

Title: US-09-807-949a-108  
Sequence: 1 GETREPL 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues  
Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	33	89.2	604	1 Y426_HUMAN	Q43309 homo sapien
2	32	86.5	781	1 GYRB_NEIGO	P22118 neiseeria g
3	31	83.8	274	1 YNP5_YEAST	P53899 saccharomyc
4	31	83.8	381	1 MRP_MYCTU	O33225 mycobacteri
5	31	83.8	406	1 DFP_ECOLI	P42385 escherichia
6	30	81.1	88	1 SNN_HUMAN	C75324 homo sapien
7	30	81.1	88	1 SNN_MOUSE	O88369 mus musculu
8	30	81.1	133	1 R8B_SYNY3	P73307 synchocyst
9	30	81.1	430	1 TRLT_HUMAN	O96924 homo sapien
10	30	81.1	430	1 TRLT_MACFA	O90932 macaca fasc
11	30	81.1	502	1 DHAB_ATPHO	P42787 atirplex ho
12	30	81.1	887	1 ORP3_HUMAN	O94415 homo sapien
13	30	81.1	3587	1 SRP1_BACSU	P27206 bacillus su
14	29	78.4	554	1 NTP3_TOBAC	P29182 nicotiana t
15	29	78.4	563	1 YMOO_YEAST	O03103 saccharomyc
16	29	78.4	709	1 WHIT_ANCAL	O16928 anopheles a
17	29	78.4	722	1 CMAA_BACTU	P66957 bacillus th
18	29	78.4	724	1 PURL_METKA	O82499 methanopyru
19	29	78.4	1263	1 RPOB_THEMA	P29338 thermotoga
20	28	75.7	140	1 POP7_HUMAN	C75817 homo sapien
21	28	75.7	200	1 PHNB_PSBAB	P09786 pseudomonas
22	28	75.7	236	1 Y4YO_RHISM	P55725 rhizobium s
23	28	75.7	311	1 MPE2_RHIME	O52292 rhizobium m
24	28	75.7	340	1 CELR_THETFU	O87550 thermomonos
25	28	75.7	340	1 RIR2_ARATH	P30651 arabidopsis
26	28	75.7	351	1 VB19_VAVAV	P33795 variola vir
27	28	75.7	385	1 HUNB_BOMMO	O18326 bombyx mori
28	28	75.7	416	1 GLVA_HELPY	O92367 helicobacte
29	28	75.7	416	1 GLVA_HELPY	P66089 helicobacte
30	28	75.7	418	1 DFP_MYCTU	P71651 mycobacteri
31	28	75.7	450	1 NCAP_RABVA	P15137 rabies viru
32	28	75.7	450	1 NCAP_RABVP	O08334 rabies viru
33	28	75.7	450	1 NCAP_RABVP	P06025 rabies viru

34	28	75.7	450	1 NCAP_RABVS	P16285 rabies viru
35	28	75.7	450	1 NCAP_RABVU	O09110 rabies viru
36	28	75.7	462	1 TRPE_THETFU	P05378 thermus the
37	28	75.7	497	1 DHAB_SPTOL	P17202 spinacia ol
38	28	75.7	500	1 DHAB_BETVU	P28237 beta vulgar
39	28	75.7	501	1 DHAB_AMAHP	O04895 amaranthus
40	28	75.7	501	1 DHAB_ARATH	O98795 arabidopsis
41	28	75.7	565	1 UBPN_HUMAN	O9680 homo sapien
42	28	75.7	566	1 UBPN_MOUSE	O96216 mus musculu
43	28	75.7	596	1 NIRS_PARDE	O51700 paracoccus
44	28	75.7	607	1 Y427_SCHPO	O09598 schizosacch
45	28	75.7	791	1 ALG2_MOUSE	O88512 mus musculu

## ALIGNMENTS

RESULT 1  
Y426\_HUMAN STANDARD; PRT; 604 AA.  
AC Q43309; Q43724;  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Hypothetical zinc finger protein KIAA0426.  
GN KIAA0426.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain;  
RX MEDLINE=98116655; PubMed=9455477;  
RA Ishikawa K.-I., Nagase T., Nakajima D., Seki N., Ohira M.,  
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;  
RT "Prediction of the coding sequences of unidentified human genes. VIII.  
RT 78 new cDNA clones from brain which code for large proteins in  
RT vitro."  
RL DNA Res. 4:307-313(1997).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Tubby B.;  
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: MAY BE INVOLVED IN TRANSCRIPTIONAL REGULATION.  
CC -!- SUBCELLULAR LOCATION: Nuclear (potential).  
CC -!- SIMILARITY: BELONGS TO THE KRUEPPL FAMILY OF C2H2-TYPE ZINC-  
CC FINGER PROTEINS.  
CC -!- SIMILARITY: CONTAINS 1 SCAN BOX.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC -----  
CC EMBL; AB007886; BAA24856.1; -  
CC EMBL; Z98745; CAB11428.2; -  
CC HSSP; P07248; 2ADR.  
CC InterPro; IPR003109; Treg SCAN.  
CC InterPro; IPR000822; Znf\_C2H2.  
CC Pfam; PF00096; Zf-C2H2; 10.  
CC Pfam; PF02023; SCAN; 1.  
CC PRINTS; PR00048; ZINCFINGER.  
CC Prodom; PD000003; Znf\_C2H2; 8.  
CC SMART; SM00431; LRR; 1.  
CC SMART; SM00355; ZNF\_C2H2; 10.  
CC PROSITE; PS50804; SCAN\_BOX; 1.  
CC PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 9.  
CC PROSITE; PS50157; ZINC\_FINGER\_C2H2\_2; 11.  
CC Hypothetical protein; Transcription regulation; DNA-binding;

```

KM Zinc-finger; Metal-binding; Nuclear protein; Repeat.
PT DOMAIN 46 128 ZINC BOX.
PT 274 512 ZINC FINGERS.
FT 274 296 C2H2-TYPE.
FT 302 324 C2H2-TYPE.
FT 330 352 C2H2-TYPE.
FT 358 380 C2H2-TYPE.
FT 386 408 C2H2-TYPE.
FT 414 436 C2H2-TYPE.
FT 442 463 C2H2-TYPE.
FT 469 491 C2H2-TYPE.
FT 497 519 C2H2-TYPE.
FT 525 547 C2H2-TYPE.
SQ SEQUENCE 604 AA; 70222 MW; 6A7F8CB515A7A73 CRC64;

Query Match 89.2%; Score 33; DB 1; Length 604;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GETREP 6
Db 224 GETREP 229

RESULT 2
CYRB_NEIGO STANDARD; PRT; 781 AA.
ID CYRB_NEIGO
AC P2218;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DNA gyrase subunit B (EC 5.99.1.3).
GN CYRB.
OS Neisseria gonorrhoeae.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=485;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MR302 / MUG116;
RX MEDLINE=9129684; PubMed=1906260;
RA Stein D.C., Danaher R.J., Cook T.M.;
RT "Characterization of a gyrase mutation responsible for low-level
RT nalidixic acid resistance in Neisseria gonorrhoeae.";
RL Antimicrob. Agents Chemother. 35:622-626(1991).
CC -1- FUNCTION: DNA GYRASE NEGATIVELY SUPERCOILS CLOSED CIRCULAR DOUBLE-
CC STRANDED DNA IN AN ATP-DEPENDENT MANNER AND ALSO CATALYZES THE
CC INTERCONVERSION OF OTHER TOPOLOGICAL ISOMERS OF DOUBLE-STRANDED
CC DNA RINGS, INCLUDING CATENANES AND KNOTTED RINGS.
CC -1- CATALYTIC ACTIVITY: ATP-dependent breakage, passage and rejoining
CC of double-stranded DNA.
CC -1- SUBUNIT: MADE UP OF TWO CHAINS. THE A CHAIN IS RESPONSIBLE FOR DNA
CC BREAKAGE AND REJOINING; THE B CHAIN CATALYZES ATP HYDROLYSIS. THE
CC ENZYME FORMS AN A2B2 TETRAMER.
CC -1- SIMILARITY: BELONGS TO THE TYPE II TOPOISOMERASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; M59981; AAA8327.1; -.
CC HSSP; P06982; 1A06.
CC InterPro; IPR003594; ATPbind_ATPase.
CC InterPro; IPR002288; DNA_gyraseB_C.
CC InterPro; IPR001241; DNA_topoisomI.
CC InterPro; IPR002836; DNAPrim_toprim.
CC InterPro; IPR004359; HIS_KIN_619.
CC Pfam; PF00204; DNA_gyraseB_1.
CC Pfam; PF00986; DNA_gyraseB_C_1.
CC Pfam; PF01751; Toprim; 1.

```

```

DR Pfam; PF02518; HATPase_C; 1.
DR PRINTS; PR00418; TP2FAMILY.
DR ProDom; PD000616; DNA_topoisomI; 1.
DR ProDom; PD149633; DNA_gyraseB_C; 1.
DR SMART; SM00387; HATPase_C; 1.
DR SMART; SM00433; TOP2c; 1.
DR TIGRFAMs; TIGR01059; gyrb; 1.
DR PROSITE; PS00177; TOPOISOMERASE_II; 1.
KM Topoisomerase; isomerase; ATP-binding.
SQ SEQUENCE 781 AA; 86516 MW; 8F9F8E69F8B258E CRC64;

Query Match 86.5%; Score 32; DB 1; Length 781;
Best Local Similarity 85.7%; Pred. No. 25;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GETREP 7
Db 152 GETREP 158

RESULT 3
YNP5_YEAST STANDARD; PRT; 274 AA.
ID YNP5_YEAST
AC P53899;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Hypothetical 31.5 kDa protein in YGP1-YCK2 intergenic region.
GN YNL155W OR N1751.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / FY1679;
RX MEDLINE=96287653; PubMed=8686380;
RA Naef F., Becam A.-M., Herbert C.J.;
RT "The sequence of 36.8 kb from the left arm of chromosome XIV reveals
RT 24 complete open reading frames: 18 correspond to new genes, one of
RT which encodes a protein similar to the human myotonic dystrophy
RT kinase."
RL Yeast 12:169-175(1996).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; X92517; CAAG3284.1; -.
CC EMBL; X71431; CAAG6042.1; -.
CC SGD; S0005099; YNL155W.
CC InterPro; IPR000058; Znf_AN1.
DR Pfam; PF01428; zf-AN1; 1.
DR SMART; SM00154; Znf_AN1; 1.
KM Hypothetical protein.
SQ SEQUENCE 274 AA; 31516 MW; 1BC3E0D932AC1365 CRC64;

Query Match 83.8%; Score 31; DB 1; Length 274;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GETREP 7
Db 98 GETREP 103

RESULT 4
MRP_MYCTU STANDARD; PRT; 381 AA.
ID MRP_MYCTU

```

```

AC 033225;
DT 30-MAY-2000 (Rel. 39, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE MTP protein homolog.
GN MRP OR RV1229C OR MT1267 OR MTC61.12C OR MTW006.01C.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteriales; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=H37Rv;
RX MEDLINE=9825987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Baaham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares R.,
RA Suleron J.E., Taylor K., Whitehead S., Barrrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
RN (2)
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Hatt D., Hickey E.,
RA Kolonyak J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains."
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: NOT KNOWN.
CC -1- SIMILARITY: BELONGS TO THE MRP/NBP35 FAMILY OF ATP-BINDING
CC PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z98260; CAB10925.1; -.
DR EMBL; AE007003; AAK45525.1; ALT_INIT.
DR TIGR; MT1267; -.
DR Tuberculist; RV1229C; -.
DR InterPro; IPR002744; DUF59.
DR InterPro; IPR000808; MTP_family.
DR Pfam; PF01883; DUF59; 1.
DR ProDom; PD005595; DUF59; 1.
DR PROSITE; PS01215; MRP; 1.
KM ATP-binding; Complete proteome.
PT NP_BIND 125 132 ATP (POTENTIAL).
SQ SEQUENCE 381 AA; 40084 MW; 4EBB63B79D6B91A CRC64;

Query Match 83.8%; Score 31; DB 1; Length 381;
Best Local Similarity 71.4%; Pred. No. 19;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 GETREPL 7
Db 101 GDTREPL 107

RESULT 5
DFF_ECOLI STANDARD; PRT; 406 AA.
ID_DFF_ECOLI

```

```

AC P24285; P76718;
DT 01-MAR-1992 (Rel. 21, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DNA/pantothenate metabolism flavoprotein.
GN DFP OR B3639.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=KL2 / MG1655;
RX MEDLINE=93315143; PubMed=7686882;
RA Burdand V.D., Plunkett G. III, Daniels D.L., Blattner F.R.;
RT "DNA sequence and analysis of 136 kilobases of the Escherichia coli
RT genome: organizational symmetry around the origin of replication."
RL Genomics 16:551-561(1993).
RN (2)
RP SEQUENCE OF 288-406 FROM N.A.
RC STRAIN=KL2;
RX MEDLINE=84057777; PubMed=6139280;
RA Lundberg L.G., Thoreason H.O., Karlstrom O.H., Nyman P.O.;
RT "Nucleotide sequence of the structural gene for dnpase of
RT Escherichia coli K-12."
RL EMBO J. 2:967-971(1983).
CC -1- FUNCTION: FLAVOPROTEIN AFFECTING SYNTHESIS OF DNA AND PANTOTHENATE
CC METABOLISM.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; LA0328; AAA61992.1; ALT_INIT.
DR EMBL; AE000441; AAC76663.1; ALT_INIT.
DR EMBL; V01578; -. NOT_ANNOTATED_CDS.
DR HSSP; Q9SWB5; 1E20.
DR EcoGene; EG10004; dfp.
DR InterPro; IPR005252; Cone_hypoth521.
DR InterPro; IPR003382; Flavoprotein.
DR Pfam; PF02441; Flavoprotein; 1.
DR TIGRfam; TIGR00521; dfp; 1.
KM Flavoprotein; Complete proteome.
PT NP_BIND 406 AA; 43438 MW; CBD11B9347E8C6AB CRC64;
SQ SEQUENCE 406 AA; 43438 MW; CBD11B9347E8C6AB CRC64;

Query Match 83.8%; Score 31; DB 1; Length 406;
Best Local Similarity 85.7%; Pred. No. 20;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GETREPL 7
Db 196 GDTREPL 202

RESULT 6
ID_SNN_HUMAN SNN_HUMAN STANDARD; PRT; 88 AA.
AC 075324;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Stanln (Ag8_1).
GN SNN.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A.

```

```

RC TISSUE=Brain;
RX MEDLINE=98325608; PubMed=9657854;
RA Dejneka N.S., Polavarapu R., Deng X., Martin-Deleon P.A.,
RA Billingsley M.L.;
RT "Chromosomal localization and characterization of the stannin (Snn)
RL gene.";
RM Mamm. Genome 9:556-564(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Endothelial cells;
RX MEDLINE=99252096; PubMed=10233894;
RA Horrovoets A.J.G., Fontijn R.D., van Zonneveld A.J., de Vries C.J.M.,
RA ten Cate J.W., Pannekoek H.;
RT "Vascular endothelial genes that are responsive to tumor necrosis
RT factor-alpha in vitro are expressed in atherosclerotic lesions,
RT including inhibitor of apoptosis protein-1, stannin, and two novel
RT genes.";
RL Blood 93:3418-3431(1999).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Amygdala;
RA Ambrose W., Winkner U., Mewes H.-W., Well B., Wiemann S.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: PLAYS A ROLE IN THE TOXIC EFFECTS OF ORGANOTINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, AF030196; AAC8427.1; -
DR EMBL, AF070673; AAC8323.1; -
DR EMBL, AL161976; CAB82314.1; -
DR GenBank, HGNC:11149; SNN.
DR MIM: 603032; -
KM Transmembrane.
FT TRANSMEM 11 31 POTENTIAL.
SQ SEQUENCE 88 AA; 9497 MW; 1E8DA73323C5D6DF CRC64;

Query Match 81.1%; Score 30; DB 1; Length 88;
Best Local Similarity 83.3%; Pred. No. 6;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

CY 1 GETREP 6
DB 54 GETREP 59

RESULT 7
SNN_MOUSE STANDARD; PRT; 88 AA.
ID SNN_MOUSE
AC O88369;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Stannin.
GN SNN.
OS Mus musculus (Mouse), and
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCB1_TaxID=10090.10116;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=Mouse;
RX MEDLINE=98325608; PubMed=9657854;
RA Dejneka N.S., Polavarapu R., Deng X., Martin-Deleon P.A.,
RA Billingsley M.L.;
RT "Chromosomal localization and characterization of the stannin (Snn)
RT gene.";

```

```

RL Mamm. Genome 9:556-564(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=Rat;
RX MEDLINE=92342135; PubMed=1635553;
RA Togga S.M., Krazy J.K., Billingsley M.L.;
RT "Molecular neurotoxicology of trimethyltin: identification of stannin,
RT a novel protein expressed in trimethyltin-sensitive cells.";
RL Mol. Pharmacol. 42:44-56(1992).
RN [3]
RP TISSUE SPECIFICITY.
RC SPECIES=Rat;
RX MEDLINE=98075763; PubMed=9413842;
RA Dejneka N.S., Patanow C.M., Polavarapu R., Togga S.M., Krazy J.K.,
RA Billingsley M.L.;
RT "Localization and characterization of stannin: relationship to
RT cellular sensitivity to organotin compounds.";
RL Neurochem. Int. 31:801-815(1997).
CC -1- FUNCTION: PLAYS A ROLE IN THE TOXIC EFFECTS OF ORGANOTINS.
CC -1- TISSUE SPECIFICITY: HIGH LEVEL OF EXPRESSION IN SPLEEN, FOLLOWED
CC BY BRAIN AND KIDNEY.
CC -1- INDUCTION: BY TRIMETHYLTIN (TMT), A TRIALKYL TIN COMPOUND WHICH IS
CC A POTENT NEUROTOXIC AGENT THAT SELECTIVELY DAMAGES SPECIFIC BRAIN
CC REGIONS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, AF030522; AAC8428.1; -
DR EMBL, M81639; -; NOT_ANNOTATED_CDS.
DR M81; MG1:1276549; Snn.
KM Transmembrane.
FT TRANSMEM 11 31 POTENTIAL.
SQ SEQUENCE 88 AA; 9501 MW; E88DA73323D816C5 CRC64;

Query Match 81.1%; Score 30; DB 1; Length 88;
Best Local Similarity 83.3%; Pred. No. 6;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

CY 1 GETREP 6
DB 54 GETREP 59

RESULT 8
RS8_SYNY3 STANDARD; PRT; 133 AA.
ID RS8_SYNY3
AC P73307;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 30S ribosomal protein S8.
GN RPS8 OR RPS8 OR SL1809.
OS Synecocystis sp. (Strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synecocystis.
OX NCB1_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneo T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hirose M., Sugitara M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Natsu K.,
RA Okumura S., Shimo S., Takeuchi C., Wada T., Watanabe A.,
RA Yamada M., Yasuda M., Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synecocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).

```

```

CC - FUNCTION: BINDS DIRECTLY TO THE CENTRAL DOMAIN OF 16S RIBOSOMAL
CC RNA (BY SIMILARITY).
CC - SIMILARITY: BELONGS TO THE S8P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D90905; BAA17336.1; -
CC HSSP: P56209; ISEI.
CC InterPro: IPR001230; Prenyl site.
CC InterPro: IPR000630; Ribosomal_S8.
CC Pfam: PF00410; Ribosomal_S8; 1.
CC ProDom: PD001098; Ribosomal_S8; 1.
CC PROSITE: PS00053; RIBOSOMAL_S8; 1.
CC Ribosomal protein; rRNA-binding; Complete proteome.
CC KW SEQUENCE 133 AA; 14666 MW; 399375HDC9489A5 CRC64;
SQ
Query Match 81.1%; Score 30; DB 1; Length 133;
Best Local Similarity 71.4%; Pred. No. 9.6;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 1 GETREPL 7
DB 68 GKTROP 74

```

```

RC TISSUE=Spleen;
RA Jikuya H., Takano J., Nomura N., Kikuno R., Nagase T., Ohara O.;
RT "The nucleotide sequence of a long cDNA clone isolated from human
RT spleen."
RT Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
RL
CC - FUNCTION: Mediates activation of NF-kappa-B. May play a role in T-
CC cell activation.
CC - SUBUNIT: Associates with TRAF1.
CC - SUBCELLULAR LOCATION: Type I membrane protein (Probable).
CC - TISSUE SPECIFICITY: Highest levels are in spleen, lymph node,
CC thymus, peripheral blood leukocytes, bone marrow and fetal liver.
CC In brain, kidney and pancreas.
CC - SIMILARITY: CONTAINS 1 TNFR-CYS REPEAT.
CC - CAUTION: Ref.4 sequence differs from that shown due to several
CC framehifts.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF319553; AAK77356.1; -
CC DR EMBL: AK027899; BAB55441.1; -
CC DR EMBL: BC001812; AAH01812.1; -
CC DR EMBL: BC017279; AAH17279.1; -
CC DR EMBL: AK074128; BAB84954.1; -
CC DR Genew: HGNC:13764; TNFRSF19L.
CC DR PROSITE: PS00652; TNFR_NGFR_1; FALSE NEG.
CC DR PROSITE: PS00050; TNFR_NGFR_2; FALSE NEG.
CC KW Receptor; Transmembrane; Glycoprotein; Signal.
CC FT SIGNAL 1 26
CC FT CHAIN 27 430
CC FT DOMAIN 27 162
CC FT TRANSMEM 163 183
CC FT FT 184 430
CC FT DOMAIN 184 430
CC FT REPEAT 50 90
CC FT DISULFID 51 65
CC FT CARBOHYD 71 90
CC FT CONFLICT 149 149 N-LINKED (GICMNC. .) (POTENTIAL).
CC FT CONFLICT 122 122 D -> S (IN REF. 4).
CC FT CONFLICT 187 187 K -> E (IN REF. 2).
CC FT CONFLICT 273 273 H -> R (IN REF. 2).
CC FT CONFLICT 379 380 DL -> TR (IN REF. 3; AAH01812).
CC SQ SEQUENCE 430 AA; 46092 MW; 4A5AB9AE3D36101 CRC64;
SQ
Query Match 81.1%; Score 30; DB 1; Length 430;
Best Local Similarity 83.3%; Pred. No. 37;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 GETREPL 6
DB 142 GKTROP 147

```

```

RESULT 10
TRLT MACFA
ID TRLT MACFA STANDARD; PRT; 430 AA.
AC 09N052;
RT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
RT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tumor necrosis factor receptor superfamily member TNFRSF19L precursor
DE (Receptor expressed in lymphoid tissues).
GN TNFRSF19L OR RELT.
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecinae; Macaca.

```

```

OX NCBI_TaxID=9541;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=21458551; PubMed=11574149;
RA Oeda N., Hida M., Kusunada J., Tanuma R., Iseki K., Hirata M., Suto Y.,
  Hirai M., Terao K., Suzuki Y., Sugano S., Hashimoto K., Kusunada J.;
RT "Assignment of 118 novel cDNAs of cynomolgus monkey brain to human
  chromosomes.";
  Gene 275:31-37(2001).
CC -1- FUNCTION: Mediates activation of NF-kappa-B (By similarity). May
  play a role in T-cell activation.
CC -1- SUBUNIT: Associates with TRAF1 (By similarity).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein (Probable).
CC -1- SIMILARITY: CONTAINS 1 TNFR-CYS REPEAT.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
-----
CC EMBL; AB046039; BAB01621.1; -.
DR InterPro; IPR001368; TNFR_C6.
DR PROSITE; PS00652; TNFR_NGFR_1; FALSE_NEG.
DR PROSITE; PS50050; TNFR_NGFR_2; FALSE_NEG.
DR SMART; SM00208; TNFR; 1.
KW Receptor; Transmembrane; Glycoprotein; Signal.
FT SIGNAL 1 26 BY SIMILARITY.
FT CHAIN 27 430 TUMOR NECROSIS FACTOR RECEPTOR
  SUPERFAMILY MEMBER TNFRSF10.
FT DOMAIN 27 162 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 163 183 POTENTIAL.
FT DOMAIN 184 430 CYTOPLASMIC (POTENTIAL).
FT REPEAT 50 90 TNFR-CYS.
FT DISULFID 51 65 BY SIMILARITY.
FT DISULFID 71 90 BY SIMILARITY.
FT CARBOHYD 149 149 N-LINKED (GLYCANC... ) (POTENTIAL).
SQ SEQUENCE 430 AA; 45850 MW; BA6DE925931E1859 CRC64;

Query March 81.1%; Score 30; DB 1; Length 430;
Best Local Similarity 83.3%; Pred. No. 37;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GETREP 6
  |||||
Db 142 GETROP 147

RESULT 11
DHAB_ATTRHO STANDARD; PRT; 502 AA.
AC P42757;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Betaine-aldehyde dehydrogenase, chloroplast precursor (EC 1.2.1.8)
  (BADH).
OS Atriplex hortensis (Mountain spinach).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
  OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
  OC Caryophyllidae; Caryophyllales; Chenopodiaceae; Atriplex.
OX NCBI_TaxID=34272;
  [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Leaf, and Stem;
RA Xiao G., Zhang G., Liu F., Chen S.;
RT "cDNA and partial genomic DNA sequence of Mountain spinach (Atriplex
  hortensis) betaine aldehyde dehydrogenase (BADH).";
  Chin. Sci. Bull. 40:741-745(1995).
RL -1- CATALYTIC ACTIVITY: Betaine aldehyde + NAD(+) + H(2)O = betaine +
  ..

```

```

CC NADH.
CC -1- PATHWAY: Betaine biosynthesis; last step.
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Chloroplast.
CC -1- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
-----
CC EMBL; X69770; CAA49425.1; -.
DR EMBL; X69772; CAA49427.1; -.
DR HSSP; P51977; 1BX5.
DR InterPro; IPR002086; Aldehyde_dehydr.
DR Pfam; PF00171; aldedh; 1.
DR PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; 1.
DR PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; 1.
KW Oxidoreductase; NAD; Chloroplast; Transit peptide.
FT TRANSIT 1 7 CHLOROPLAST (POTENTIAL).
FT CHAIN 8 502 BETAIN-ALDEHYDE DEHYDROGENASE.
FT NP_BIND 240 245 MAD (ADP PART) (BY SIMILARITY).
FT ACT_SITE 262 262 BY SIMILARITY.
FT ACT_SITE 296 296 BY SIMILARITY.
SQ SEQUENCE 502 AA; 55270 MW; 54F1A05EB9C4281E CRC64;

Query March 81.1%; Score 30; DB 1; Length 502;
Best Local Similarity 85.7%; Pred. No. 44;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GETREP 7
  |||||
Db 14 GEMREPL 20

RESULT 12
ORP3_HUMAN STANDARD; PRT; 887 AA.
AC Q9H4I5; Q9B2F4; O14591; Q9UED6; O43357; O43358;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Oysterol binding protein-related protein 3 (OSBP-related protein 3)
  (ORP-3).
DE (ORP-3).
GN OSBP3 OR ORP3 OR KIAA0704.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
  [1]
RP SEQUENCE FROM N.A.
RA Gregorio-King C.C., Collier G.C., McMillan J., Waugh C.,
  Kirkland M.A.;
RT "ORP3: a novel human oysterol-binding protein gene expressed in CD34+
  stem/progenitor cells.";
  Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21601154; PubMed=11735225;
RA Jaworski C.J., Moreira E., Li A., Lee R., Rodriguez I.R.;
RT "A family of 12 human genes containing oysterol-binding domains.";
  Genomics 78:185-196(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Uterus;
RA Straube R.;
RT Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 1-245 FROM N.A.
RX MEDLINE=21376257; PubMed=11483621;

```



RA Lehto M., Laitinen S., Chinetti G., Johansson M., Enholm C.,  
 RA Staelen B., Ikonen E., Oikonen V.M.;  
 RT "The OSBP-related protein family in humans."  
 RL J. Lipid Res. 42:1203-1213(2001).  
 RN [5]  
 RP SEQUENCE OF 221-887 FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=98403880; PubMed=9734811;  
 RA Ichikawa K.-I., Nagase T., Suyama M., Miyajima N., Tanaka A.,  
 RA Kotani H., Nomura N., Ohara O.,  
 RT "Prediction of the coding sequences of unidentified human genes. X.  
 RT The complete sequences of 100 new cDNA clones from brain which can  
 RT code for large proteins in vitro."  
 RL DNA Res. 5:169-176(1998).  
 RN [6]  
 RP SEQUENCE OF 291-467 FROM N.A.  
 RA Dante M., Kellen J., O'Brian D.;  
 RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.  
 RN [7]  
 RP SEQUENCE OF 500-649 FROM N.A.  
 RA Dante M., Kellen J., O'Brian D.;  
 RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.  
 RN [8]  
 RP SEQUENCE OF 676-887 FROM N.A.  
 RA Dauphin S., Mooney A.;  
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: BELONGS TO THE OSBP FAMILY.  
 CC -1- SIMILARITY: CONTAINS 1 PH DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AY008372; AAG23400.1; -;  
 DR EMBL: AF392444; AAL0657.1; -;  
 DR EMBL: BC017731; AAH17731.1; -;  
 DR EMBL: AB014604; BAA31679.1; -;  
 DR EMBL: AF323727; AAG53408.1; -;  
 DR EMBL: AC003093; AAB83939.1; -;  
 DR EMBL: AC004016; AAC26986.1; -;  
 DR EMBL: AC004016; AAC26985.1; -;  
 DR GeneW; HGNC:15370; OSBP.L3.  
 DR MIM; 606732; -;  
 DR InterPro; IPR000648; Oxyterol\_BP.  
 DR InterPro; IPR001849; PH.  
 DR Pfam; PF00169; PH.1.  
 DR Pfam; PF01237; Oxyterol\_BP.1.  
 DR SMART; SM00233; PH.1.  
 DR PROSITE; PS01013; OSBP.1.  
 DR PROSITE; PS5003; PH\_DOMAIN.1.  
 DR KMW lipid transport; Transport.  
 FT DOMAIN 51 146 PH.  
 FT CONFLICT 676 676 V->S (IN REF. 8).  
 SQ SEQUENCE 887 AA; 101223 MW; 12E16912B3F2B99 CRC64;

Query Match 81.1%; Score 30; DB 1; Length 887;  
 Best Local Similarity 85.7%; Pred. No. 83;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GETREPL 7  
 DB 205 GETREPL 211

RESULT 13  
 SRFL\_BACSU STANDARD; PRT; 3587 AA.  
 ID SRFL\_BACSU STANDARD; PRT; 3587 AA.  
 AC P27206;  
 DT 01-AUG-1992 (Rel. 23, Created)

DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Surfactin synthetase subunit 1.  
 GN SRFPA OR SRFPA1 OR SRFPA.  
 OS Bacillus subtilis.  
 OC Bacillus, Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OX NCBI\_TaxId=1423;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE=93181186; PubMed=8441623;  
 RA Fuma S., Fujishima Y., Corbelli N., D'Souza C., Nakano M.M.,  
 RA Zuber P., Yamane K.;  
 RT "Nucleotide sequence of 5' portion of srfA that contains the region  
 RT required for competence establishment in Bacillus subtilis."  
 RL Nucleic Acids Res. 21:93-97(1993).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168 / JH642;  
 RX MEDLINE=93360813; PubMed=8355609;  
 RA Cosmina P., Rodriguez F., de Ferra F., Grandi G., Perago M.,  
 RA Venema G., van Sinderen D.;  
 RT "Sequence and analysis of the genetic locus responsible for surfactin  
 RT synthesis in Bacillus subtilis."  
 RL Mol. Microbiol. 8:821-831(1993).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE=97124189; PubMed=8969502;  
 RA Yamane K., Kumano M., Kurita K.;  
 RT "The 25 degrees-36 degrees region of the Bacillus subtilis  
 RT chromosome: determination of the sequence of a 146 kb segment and  
 RT identification of 113 genes."  
 RL Microbiology 142:3047-3056(1996).  
 RN [4]  
 RP SEQUENCE OF 1-460 FROM N.A.  
 RX MEDLINE=91154134; PubMed=1847909;  
 RA Nakano M.M., Magnusson R., Myers A.M., Curry J., Groesman A.D.,  
 RA Zuber P.;  
 RT "srfA is an operon required for surfactin production, competence  
 RT development, and efficient sporulation in Bacillus subtilis."  
 RL J. Bacteriol. 173:1770-1778(1991).  
 RN [5]  
 RP SEQUENCE OF 1-38 FROM N.A.  
 RX MEDLINE=91358326; PubMed=1715856;  
 RA Nakano M.M., Xia L., Zuber P.;  
 RT "Transcription initiation region of the srfA operon, which is  
 RT controlled by the comp-cpmA signal transduction system in Bacillus  
 RT subtilis."  
 RL J. Bacteriol. 173:5487-5493(1991).  
 RN [6]  
 RP SEQUENCE OF 1-64 FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE=95219080; PubMed=7704255;  
 RA Fujishima Y., Yamane K.;  
 RT "A 10 kb nucleotide sequence at the 5' flanking region (32 degrees)  
 RT of srfA of the Bacillus subtilis chromosome."  
 RL Microbiology 141:277-279(1995).  
 CC -1- FUNCTION: THIS PROTEIN IS A MULTIFUNCTIONAL ENZYME ABLE TO  
 CC ACTIVATE AND POLYMERIZE THE AMINO ACIDS LEU, GLU, ASP AND VAL.  
 CC ACTIVATION SITES FOR THESE AA CONSIST OF INDIVIDUAL DOMAINS.  
 CC -1- COPFACTOR: CONTAINS 3 COVALENTLY BOUND PHOSPHOPANTHEINES.  
 CC -1- PATHWAY: Cyclic peptide antibiotic surfactin biosynthesis.  
 CC -1- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME  
 CC FAMILY.  
 CC -1- SIMILARITY: CONTAINS 3 ACYL CARRIER DOMAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

```

CC or send an email to license@ib-sib.ch).
CC -----
DR EMBL; D13262; BAA02522.1; -
DR EMBL; X70356; CAA49816.1; -
DR EMBL; D50453; BAA06882.1; -
DR EMBL; M59939; AAA22815.1; -
DR EMBL; M64702; AAA22816.1; -
DR EMBL; D30762; BAA21034.1; -
DR EMBL; Z99105; CAB12142.1; -
DR PIR; S35517; S35517.
DR HSSP; P14687; 1AMU.
DR SUDLisc; B610168; srFAA.
DR InterPro; IPR000873; AMP-bind.
DR InterPro; IPR001242; Condensatn.
DR InterPro; IPR003880; Pantene attach.
DR Pfam; PF00550; AMP-binding; 3.
DR Pfam; PF00668; Condensation; 4.
DR PRINTS; PR00154; AMPBINDING.
DR PROSITE; PS00012; PHOSPHOPANTHEINE; 3.
DR PROSITE; PS00455; AMP_BINDING; 3.
DR PROSITE; PS00075; ACP_DOMAIN; 3.
KW Ligase; Anticlotic biosynthesis; Phosphopantetheine; Sporulation;
KW Multifunctional enzyme; Repeat; Complete proteome.
FT REPEAT ? 1047 DOMAIN 1 (GLU-ACTIVATING).
FT REPEAT ? 2086 DOMAIN 2 (LEU-ACTIVATING).
FT REPEAT ? 3114 DOMAIN 3 (D-LEU-ACTIVATING).
FT DOMAIN 976 1043 ACYL CARRIER (ACP) 1.
FT DOMAIN 2015 2082 ACYL CARRIER (ACP) 2.
FT DOMAIN 3043 3109 ACYL CARRIER (ACP) 3.
FT BINDING 1006 1006 PHOSPHOPANTHEINE (POTENTIAL).
FT BINDING 2045 2045 PHOSPHOPANTHEINE (POTENTIAL).
FT BINDING 3073 3073 PHOSPHOPANTHEINE (POTENTIAL).
FT BINDING 146 146 I -> M (IN REF. 2 AND 3).
FT CONFLICT 151 151 L -> V (IN REF. 2 AND 3).
FT CONFLICT 165 165 A -> G (IN REF. 2 AND 4).
FT CONFLICT 281 281 T -> Q (IN REF. 2 AND 3).
FT CONFLICT 460 460 D -> T (IN REF. 3 AND 3).
FT CONFLICT 540 540 P -> A (IN REF. 2 AND 3).
FT CONFLICT 562 562 I -> Y (IN REF. 2 AND 3).
FT CONFLICT 639 640 GS -> PT (IN REF. 2 AND 3).
FT CONFLICT 644 644 R -> P (IN REF. 2 AND 3).
FT CONFLICT 647 647 L -> F (IN REF. 2 AND 3).
FT CONFLICT 649 649 A -> P (IN REF. 2 AND 3).
FT CONFLICT 1026 1026 L -> Q (IN REF. 2 AND 3).
FT CONFLICT 1165 1116 SLAORTYVSGFEDAGVGMPPAAILEGLDLOKLERAP
GGLIRHRESLR -> HHSSEHSASSRMSAIIHSSSN
SEGRYSKAGNISGINTPRVIE (IN REF. 2 AND
3).
FT CONFLICT 1132 1133 DS -> VC (IN REF. 2 AND 3).
FT CONFLICT 1164 1164 R -> A (IN REF. 2 AND 3).
FT CONFLICT 1458 1458 V -> L (IN REF. 2 AND 3).
FT CONFLICT 1850 1850 S -> T (IN REF. 2 AND 3).
FT CONFLICT 1894 1894 D -> T (IN REF. 2 AND 3).
FT CONFLICT 1911 1911 R -> A (IN REF. 2 AND 3).
FT CONFLICT 1973 1980 LRLSKWT -> PAVFIQMD (IN REF. 2 AND 3).
FT CONFLICT 2052 2052 L -> P (IN REF. 2 AND 3).
FT CONFLICT 2218 2218 H -> Q (IN REF. 2 AND 3).
FT CONFLICT 2265 2266 QQ -> HR (IN REF. 2 AND 3).
FT CONFLICT 2291 2291 V -> L (IN REF. 2 AND 3).
FT CONFLICT 2349 2349 S -> C (IN REF. 2 AND 3).
FT CONFLICT 2428 2428 R -> S (IN REF. 2 AND 3).
FT CONFLICT 2611 2612 RC -> AV (IN REF. 2 AND 3).
FT CONFLICT 2685 2685 E -> ENPE (IN REF. 2 AND 3).
FT CONFLICT 2755 2755 T -> S (IN REF. 2 AND 3).
FT CONFLICT 2896 2897 SP -> TA (IN REF. 2 AND 3).
FT CONFLICT 3025 3025 P -> N (IN REF. 2 AND 3).
FT CONFLICT 3096 3096 F -> N (IN REF. 2 AND 3).
FT CONFLICT 3271 3271 A -> S (IN REF. 2 AND 3).
FT CONFLICT 3316 3316 R -> S (IN REF. 2 AND 3).
FT CONFLICT 3451 3451 Y -> S (IN REF. 2 AND 3).
FT CONFLICT 3483 3484 DE -> HQ (IN REF. 2 AND 3).
FT CONFLICT 3487 3490 DAGL -> HPPR (IN REF. 2 AND 3).

```

```

FT CONFLICT 3493 3493 R -> T (IN REF. 2 AND 3).
FT CONFLICT 3495 3495 E -> H (IN REF. 2 AND 3).
FT CONFLICT 3499 3500 GQ -> PH (IN REF. 2 AND 3).
FT CONFLICT 3507 3507 R -> E (IN REF. 2 AND 3).
SQ SEQUENCE 3587 AA; 402424 MM; A1B2DAHF93DEEA CRC64;

Query Match 81.1%; Score 30; DB 1; Length 3587;
Best Local Similarity 85.7%; Pred. No. 4.1e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GETREPL 7
Db 2247 GETREPL 2253

RESULT 14
NTP3_TOBAC STANDARD; PRT; 554 AA.
AC P29162;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Pollen-specific protein NTP3 precursor.
OS Nicotiana tabacum (Common tobacco).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.
OX NCBI_Taxid=4097;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=cv. Petit Havana SR1; TISSUE=Pollen;
RC MEDLINE=92288302; Pubmed=1600146;
RA Weterings K., Reijnen W., van Aartsen R., Kortstee A., Spijkers J.,
RA van Herpen M., Schrauwen J., Mulleens G.;
RT "Characterization of a pollen-specific cDNA clone from Nicotiana
RT tabacum expressed during microgametogenesis and germination.";
RL Plant Mol. Biol. 18:1101-1111(1992).
CC -1- FUNCTION: PRECISE FUNCTION UNKNOWN BUT PROBABLY HAS AN IMPORTANT
CC -1- ROLE DURING GERMINATION AND/OR TUBE GROWTH.
CC -1- TISSUE SPECIFICITY: POLLEN.
CC -1- DEVELOPMENTAL STAGE: APPEAR AFTER THE FIRST HAPLOID MITOSIS AND
CC ARE EXPRESSED DURING MICROGAMETOGENSESIS, GERMINATION AND TUBE
CC GROWTH.
CC -1- INDUCTION: EXPRESSION REGULATED BY THE HAPLOID GAMETOPHYTE ITSELF.
CC -1- SIMILARITY: CONTAINS 3 PLASTOCYANIN-LIKE DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ibt-sib.ch/announce/
CC or send an email to license@ib-sib.ch).
CC -----
DR EMBL; X61146; CAA43454.1; -
DR PIR; S22495; S22495.
DR InterPro; IPR001117; Cu-oxidase.
DR Pfam; PF00394; Cu-oxidase; 1.
KW Signal; Germinalin; Glycoprotein; Repeat.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 554 POLLEN-SPECIFIC PROTEIN NTP303.
FT DOMAIN 22 143 PLASTOCYANIN-LIKE 1.
FT DOMAIN 196 296 PLASTOCYANIN-LIKE 2.
FT DOMAIN 411 521 PLASTOCYANIN-LIKE 3.
FT CARBOHYD 31 31 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 59 59 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 108 108 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 332 332 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 352 352 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 554 AA; 62033 MM; 90380AB1F52B2F85 CRC64;

```

Query Match 78.4%; Score 29; DB 1; Length 554;  
 Best Local Similarity 71.4%; Pred. No. 82;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GETREPL 7  
 |||:  
 Db 200 GEAREPL 206

## RESULT 15

YMN0\_YEAST  
 ID YMN0\_YEAST STANDARD; PRT; 563 AA.  
 AC Q03103;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical 65.0 kDa protein in COX14-COS3 intergenic region  
 DE precursor.  
 DE YML130C OR YM4987.05C.  
 GN Saccharomyces cerevisiae (Baker's Yeast).  
 OS Saccharomycetes cerevisiae (Baker's Yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288C / AB972;  
 RA Bowman S., Louis E.J., Barrell B.G., Rajandream N.A., Walsh S.V.;  
 RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; Z50178; CAA90553.1; -  
 DR SGD; S0004539; YML130C.  
 KM Hypothetical protein; Signal.  
 FT SIGNAL 1 20  
 FT CHAIN 21 563  
 FT SQ SEQUENCE 563 AA; 65032 MW; 928CE700AE6137EF CRC64;  
 FT POTENTIAL.

Query Match 78.4%; Score 29; DB 1; Length 563;  
 Best Local Similarity 83.3%; Pred. No. 84;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 ETRRPL 7  
 |||:  
 Db 551 ETRRPI 556

Search completed: January 29, 2003, 14:00:25  
 Job time : 5.84615 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2003, 13:56:21 ; Search time 19.5641 Seconds  
(without alignments)  
73.723 Million cell updates/sec

Title: US-09-807-949A-108

Perfect score: 37

Sequence: 1 GETREPL 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: SP\_ARCHAEA:\*  
2: SP\_BACTERIA:\*  
3: SP\_FUNGI:\*  
4: SP\_HUMAN:\*  
5: SP\_INVERTEBRATE:\*  
6: SP\_MAMMAL:\*  
7: SP\_MHC:\*  
8: SP\_ORGANELLE:\*  
9: SP\_PHAGE:\*  
10: SP\_PLANT:\*  
11: SP\_RODENT:\*  
12: SP\_VIRUS:\*  
13: SP\_VERTEBRATE:\*  
14: SP\_UNCLASSIFIED:\*  
15: SP\_VIRUS:\*  
16: SP\_BACTERIAP:\*  
17: SP\_ARCHAEA:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query	Match Length	DB ID	Description
1	33	89.2	244	4 Q9BR78	Q9BR78 homo sapien
2	33	89.2	613	17 Q9UZR8	Q9UZR8 pyrococcus
3	33	89.2	618	17 Q58865	Q58865 pyrococcus
4	33	89.2	793	4 Q14968	Q14968 homo sapien
5	33	89.2	963	4 Q14114	Q14114 homo sapien
6	33	89.2	1121	10 Q9M360	Q9M360 arabidopsis
7	32	86.5	431	17 Q8ZTY0	Q8ZTY0 pyrobaculum
8	32	86.5	473	17 Q9HPK0	Q9HPK0 halobacteri
9	32	86.5	602	5 Q46138	Q46138 lineus sang
10	32	86.5	611	4 Q95251	Q95251 homo sapien
11	32	86.5	736	16 Q9RYW6	Q9RYW6 deinococcus
12	32	86.5	796	16 Q9K1E4	Q9K1E4 neisseria m
13	32	86.5	949	16 Q9YOH4	Q9YOH4 neisseria m
14	31	83.8	115	5 Q9B1H5	Q9B1H5 dirosophila
15	31	83.8	144	17 Q9HRN7	Q9HRN7 anopheles g
16	31	83.8	144	17 Q9HRN7	Q9HRN7 halobacteri

17	31	83.8	168	5 Q9XXC6	Q9XXC6 caenorhabdi
18	31	83.8	325	16 Q9XA63	Q9XA63 streptomyce
19	31	83.8	331	16 Q8UK92	Q8UK92 streptomyce
20	31	83.8	398	16 Q8R9S8	Q8R9S8 thermomane
21	31	83.8	405	16 Q8ZTP4	Q8ZTP4 yesinia pe
22	31	83.8	407	16 Q8ZL47	Q8ZL47 salmoneilla
23	31	83.8	407	16 Q8ZDH3	Q8ZDH3 salmoneilla
24	31	83.8	430	16 Q8XD42	Q8XD42 escherichia
25	31	83.8	440	10 Q9ASR5	Q9ASR5 arabidopsis
26	31	83.8	443	6 Q9GKR7	Q9GKR7 macaca faec
27	31	83.8	447	10 Q6S629	Q6S629 arabidopsis
28	31	83.8	451	16 Q9XA10	Q9XA10 streptomyce
29	31	83.8	474	2 Q9AJK6	Q9AJK6 marine cfb-
30	31	83.8	474	2 Q9AJK4	Q9AJK4 marine cfb-
31	31	83.8	474	2 Q9AJJ8	Q9AJJ8 marine cfb-
32	31	83.8	474	2 Q9AJI7	Q9AJI7 marine cfb-
33	31	83.8	474	2 Q9AJI2	Q9AJI2 marine cfb-
34	31	83.8	474	2 Q9AJI1	Q9AJI1 marine cfb-
35	31	83.8	474	2 Q9AJH3	Q9AJH3 marine cfb-
36	31	83.8	612	17 Q8TNY7	Q8TNY7 methanosarc
37	31	83.8	741	5 Q9NGV3	Q9NGV3 drosophila
38	31	83.8	741	5 Q9RYR1	Q9RYR1 drosophila
39	31	83.8	851	13 Q90Z50	Q90Z50 fugu rubrip
40	31	83.8	998	17 Q96Z52	Q96Z52 sulfolobus
41	31	83.8	1013	12 Q91LX9	Q91LX9 retroperito
42	30	81.1	130	16 Q9CFM8	Q9CFM8 lactococcus
43	30	81.1	135	17 Q9Y8N6	Q9Y8N6 aetopyrum p
44	30	81.1	179	16 Q8ZRD6	Q8ZRD6 salmoneilla
45	30	81.1	202	17 Q30084	Q30084 archaeoglob

## ALIGNMENTS

RESULT 1  
ID Q9BR78 PRELIMINARY; PRT; 244 AA.

AC Q9BR78; 01-JUN-2001 (T-EMBLrel. 17, Created)  
DT 01-JUN-2001 (T-EMBLrel. 17, Last sequence update)  
DT 01-MAR-2002 (T-EMBLrel. 20, Last annotation update)  
DE Similar to low density lipoprotein receptor-related protein 8,  
DE apolipoprotein e receptor (Fragment).  
OS Homo sapiens (Human).  
OC Bacteria; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=LUNG;  
RA Strauberg R.;  
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC006443; AA06443.1; -  
KW Lipoprotein; Receptor.  
FT NON TER 1  
SQ SEQUENCE 244 AA; 2668 MW; 606EA634E25F56 CRC64;

Query Match 89.2%; Score 33; DB 4; Length 244;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GETREP 6  
DB 189 GETREP 194

RESULT 2  
ID Q9UZR8 PRELIMINARY; PRT; 613 AA.  
AC Q9UZR8; 01-MAY-2000 (T-EMBLrel. 13, Created)  
DT 01-MAY-2000 (T-EMBLrel. 13, Last sequence update)  
DT 01-MAR-2002 (T-EMBLrel. 20, Last annotation update)

```
DE Indolepyruvate ferredoxin oxidoreductase, subunit alpha (IOra-2).
GN IOra-2 OR PAB0718.
OS Pyrococcus abyssi.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=29292;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ORSAV;
RA Heilig R.;
RT "Pyrococcus abyssi genome sequence: insights into archaeal chromosome
   structure and evolution."
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ248286; CAB49988.1; -.
DR InterPro; IPR001450; 4Fe4S_ferredoxin.
DR InterPro; IPR002880; POR_N.
DR Pfam; PF00037; Ier4; 1.
DR PROSITE; PS00198; 4Fe4S_FERREDOXIN; 1.
KM Pyruvate; Complete proteome.
SQ SEQUENCE 613 AA; 67875 MW; 85FB55659C13PA8F CRC64;

Query Match          89.2%; Score 33; DB 17; Length 613;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 1 GETREP 6
Db 547 GETREP 552

RESULT 3
ID 058865 PRELIMINARY; PRT; 618 AA.
AC 058865;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE 618AA long hypothetical indolepyruvate ferredoxin oxidoreductase alpha
   subunit.
GN PH118.
OS Pyrococcus horikoshii.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=53953;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OT3;
RX MEDLINE=98344137; PubMed=9679194;
RA Kawarabayashi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
   Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
   Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohkuku Y.,
   Funahashi T., Tanaka T., Kudo Y., Yamazaki J., Kushida N., Oguchi A.,
   Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
   Maechui Y., Shizuya H., Kikuchi H.;
RA "Complete sequence and gene organization of the genome of a hyper-
   thermophilic archaeobacterium, Pyrococcus horikoshii OT3."
RT thermophilic archaeobacterium, Pyrococcus horikoshii OT3.;
RL DNA Res. 5:55-76(1998).
DR EMBL; AP000005; BAA30238.1; -.
DR HSSP; P46797; IVDM.
DR InterPro; IPR001450; 4Fe4S_ferredoxin.
DR InterPro; IPR002114; Hpr_Serp_site.
DR InterPro; IPR002880; POR_N.
DR Pfam; PF00037; Ier4; 2.
DR Pfam; PF01855; POR_N; 1.
DR PROSITE; PS00198; 4Fe4S_FERREDOXIN; 1.
DR PROSITE; PS00589; PTS_HPR_SER; UNKNOWN 1.
KM Pyruvate; Iron-sulfur; Complete proteome.
SQ SEQUENCE 618 AA; 68303 MW; D1925D15D392B15F CRC64;

Query Match          89.2%; Score 33; DB 17; Length 618;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
DE Indolepyruvate ferredoxin oxidoreductase, subunit alpha (IOra-2).
GN IOra-2 OR PAB0718.
OS Pyrococcus abyssi.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=29292;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ORSAV;
RA Heilig R.;
RT "Pyrococcus abyssi genome sequence: insights into archaeal chromosome
   structure and evolution."
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ248286; CAB49988.1; -.
DR InterPro; IPR001450; 4Fe4S_ferredoxin.
DR InterPro; IPR002880; POR_N.
DR Pfam; PF00037; Ier4; 1.
DR PROSITE; PS00198; 4Fe4S_FERREDOXIN; 1.
DR PROSITE; PS00589; PTS_HPR_SER; UNKNOWN 1.
KM Pyruvate; Iron-sulfur; Complete proteome.
SQ SEQUENCE 618 AA; 68303 MW; D1925D15D392B15F CRC64;

Query Match          89.2%; Score 33; DB 4; Length 793;
Best Local Similarity 100.0%; Pred. No. 89;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 1 GETREP 6
Db 738 GETREP 743

RESULT 4
ID 014968 PRELIMINARY; PRT; 793 AA.
AC 014968;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE APOER2delta-7.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97236806; PubMed=9079678;
RA Kim D., Magocori K., Inoue T.R., Mao C.C., Kim H., Suzuki H.,
   Fujita T., Endo Y., Saeki S., Yamamoto T.T.;
RT "Exon/Intron organization, chromosome localization, alternative
   splicing, and transcription units of the human apolipoprotein E
   receptor 2 gene."
RL J. Biol. Chem. 272:8498-8504(1997).
DR EMBL; D86407; BAA21825.1; -.
DR EMBL; D86389; BAA21825.1; JOINED.
DR EMBL; D86390; BAA21825.1; JOINED.
DR EMBL; D86391; BAA21825.1; JOINED.
DR EMBL; D86392; BAA21825.1; JOINED.
DR EMBL; D86395; BAA21825.1; JOINED.
DR EMBL; D86396; BAA21825.1; JOINED.
DR EMBL; D86397; BAA21825.1; JOINED.
DR EMBL; D86398; BAA21825.1; JOINED.
DR EMBL; D86399; BAA21825.1; JOINED.
DR EMBL; D86400; BAA21825.1; JOINED.
DR EMBL; D86401; BAA21825.1; JOINED.
DR EMBL; D86402; BAA21825.1; JOINED.
DR EMBL; D86403; BAA21825.1; JOINED.
DR EMBL; D86404; BAA21825.1; JOINED.
DR EMBL; D86405; BAA21825.1; JOINED.
DR EMBL; D86406; BAA21825.1; JOINED.
DR HSSP; Q07954; ICR8.
DR InterPro; IPR000152; Asx_hydroxyl.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR000033; Ldl_receptor_rep.
DR InterPro; IPR002172; LDL_recept_A.
DR Pfam; PF00008; EGF; 2.
DR Pfam; PF00057; Ldl_recept_a; 3.
DR Pfam; PF00058; Ldl_recept_b; 5.
DR SMART; SM00179; EGF_CA; 1.
DR SMART; SM00001; EGF_like; 2.
DR SMART; SM00192; LDLR; 3.
DR SMART; SM00335; LY; 5.
DR PROSITE; PS00010; ASX_HYDROXYL; UNKNOWN_2.
DR PROSITE; PS01186; EGF_2; UNKNOWN_2.
DR PROSITE; PS01187; EGF_CA; 1.
DR PROSITE; PS01209; LDLR_1; 3.
DR PROSITE; PS50068; LDLR_2; 3.
KM Calcium-binding; EGF-like domain; Glycoprotein; Repeat.
SQ SEQUENCE 793 AA; 88013 MW; 4B6A5EB8ED0D5571 CRC64;

Query Match          89.2%; Score 33; DB 4; Length 793;
Best Local Similarity 100.0%; Pred. No. 89;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

RESULT 5  
014114 PRELIMINARY; PRT; 963 AA.  
AC 014114;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE Apolipoprotein E receptor 2 precursor.  
OS Homo sapiens (human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=PLACENTA;  
RX MEDLINE=96215242; PubMed=8626535;  
RA Kim D.H., Iijima H., Goto K., Sakai J., Ishii H., Kim H.J., Suzuki H.,  
RA Kondo H., Saeki S., Yamamoto T.;  
RT "Human apolipoprotein E receptor 2. A novel lipoprotein receptor of  
RT the low density lipoprotein receptor family predominantly expressed in  
RT brain.";  
RL J. Biol. Chem. 271:8373-8380(1996).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=97236806; PubMed=9079678;  
RA Kim D., Maggiori K., Inoue T.R., Mao C.C., Kim H., Suzuki H.,  
RA Fujita T., Endo Y., Saeki S., Yamamoto T.T.;  
RT "Exon/intron organization, chromosome localization, alternative  
RT splicing, and transcription units of the human apolipoprotein E  
RT receptor 2 gene.";  
RL J. Biol. Chem. 272:8498-8504(1997).  
DR EMBL; D86407; BAA21824.1; JOINED.  
DR EMBL; D86389; BAA21824.1; JOINED.  
DR EMBL; D86390; BAA21824.1; JOINED.  
DR EMBL; D86391; BAA21824.1; JOINED.  
DR EMBL; D86392; BAA21824.1; JOINED.  
DR EMBL; D86393; BAA21824.1; JOINED.  
DR EMBL; D86394; BAA21824.1; JOINED.  
DR EMBL; D86395; BAA21824.1; JOINED.  
DR EMBL; D86396; BAA21824.1; JOINED.  
DR EMBL; D86397; BAA21824.1; JOINED.  
DR EMBL; D86398; BAA21824.1; JOINED.  
DR EMBL; D86399; BAA21824.1; JOINED.  
DR EMBL; D86400; BAA21824.1; JOINED.  
DR EMBL; D86401; BAA21824.1; JOINED.  
DR EMBL; D86402; BAA21824.1; JOINED.  
DR EMBL; D86403; BAA21824.1; JOINED.  
DR EMBL; D86404; BAA21824.1; JOINED.  
DR EMBL; D86405; BAA21824.1; JOINED.  
DR EMBL; D86406; BAA21824.1; JOINED.  
DR EMBL; D86407; BAA21824.1; JOINED.  
DR HSSP; P01130; 1A0J.  
DR InterPro; IPR000152; Asx\_hydroxyl.  
DR InterPro; IPR000561; EGF-like.  
DR InterPro; IPR001881; EGF Ca.  
DR InterPro; IPR000033; LDL\_receptor\_rep.  
DR InterPro; IPR002172; LDL\_recept\_A.  
DR Pfam; PF00057; LDL\_recept\_a; 7.  
DR Pfam; PF00058; LDL\_recept\_b; 5.  
DR PRINTS; PRO0261; LDLRECEPTOR.  
DR SMART; SM00179; EGF\_CA\_1.  
DR SMART; SM00001; EGF\_like; 2.  
DR SMART; SM00192; LDLA; 7.  
DR SMART; SM00135; LY; 5.  
DR PROSITE; PS00010; ASX\_HYDROXYL; UNKNOWN\_2.  
DR PROSITE; PS01186; EGF\_2; UNKNOWN\_2.  
DR PROSITE; PS01187; EGF\_CA; 1.  
DR PROSITE; PS01203; LDLA\_1; 7.  
DR PROSITE; PS50068; LDLA\_2; 7.  
KW Calcium-binding; EGF-like domain; Glycoprotein; Lipoprotein; Repeat;  
KW Signal.  
FT SIGNAL. 1 41 POTENTIAL.

FT CHAIN 42 963 APOLIPOPROTEIN E RECEPTOR 2.  
SQ SEQUENCE 963 AA; 105715 MM; 17CE2304715C1A6B CRC64;  
Query Match 89.2%; Score 33; DB 4; Length 963;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 GETREP 6  
Db 908 GETREP 913  
RESULT 6  
Q9M360 PRELIMINARY; PRT; 1121 AA.  
AC Q9M360;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)  
DE Hypoetical 127.9 kDa protein.  
GN P15G16.170.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC euroside II; Brassicales; Brassicaceae; Arabidopsie.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA De Haan M., Maarse A.C., Grivell L.A., Meves H.W., Lemcke K.,  
RA Mayer K.F.X., Queclier F., Salanoubat M.;  
RT Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
RL [2]  
RP SEQUENCE FROM N.A.  
RA EU Arabidopsis sequencing project;  
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; A1132959; CAB71108.1; -.  
KW Hypoetical protein.  
SQ SEQUENCE 1121 AA; 127920 MM; 13F4489745E7227F CRC64;  
Query Match 89.2%; Score 33; DB 10; Length 1121;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 GETREP 6  
Db 660 GETREP 665  
RESULT 7  
Q8ZYYO PRELIMINARY; PRT; 431 AA.  
AC Q8ZYYO;  
DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Amino acid permease.  
GN PAB0556.  
OS Pyrobaculum aerophilum.  
OC Archaea; Crenarchaeota; Thermoprotei; Thermoproteales;  
OC Thermoproteaceae; Pyrobaculum.  
OX NCBI\_TaxID=13773;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=IM2 / ATCC 51768 / DSM 7523;  
RX PubMed=11792869;  
RA Fitz-Gibbon S.T., Ladner H., Kim U.-J., Stetter K.O., Simon M.I.,  
RA Miller U.H.;  
RT "Genome sequence of the hyperthermophilic crenarchaeon Pyrobaculum  
RT aerophilum.";  
RL Ercoc. Natl. Acad. Sci. U.S.A. 99:984-989(2002).  
DR EMBL; AE009773; AAL62861.1; -.  
DR InterPro; IPR002293; AA/re1 permease.  
DR InterPro; IPR004841; Permease.

DR Pfam; PF00324; aa\_permeases; 1.  
 KM Complete proteome-  
 SQ SEQUENCE 431 AA; 47158 MW; ACE7A850453C78B CRC64;  
 Query Match 86.5%; Score 32; DB 17; Length 431;  
 Best Local Similarity 85.7%; Pred. No. 78;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GETREPL 7  
 Db 214 GAREPL 220

RESULT 8  
 Q9HFKO PRELIMINARY; PRT; 473 AA.  
 AC Q9HFKO;  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
 DE Glycine dehydrogenase subunit 2.  
 GN GCY2 OR YNG1601G.  
 OS Halobacterium sp. (strain NRC-1).  
 OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;  
 OC Halobacteriaceae; Halobacterium.  
 CX NCBI\_TaxID=64091;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20504483; PubMed=11016950;  
 RA Ng W.V., Kennedy S.P., Mahairas G.G., Bergquist B., Pan M.,  
 RA Shukla H.D., Laezy S.R., Baliga N.S., Thorsson V., Shogna J.,  
 RA Swartzell S., Weir D., Hall J., Dahl T.A., Weir R., Goo Y.A.,  
 RA Leitchner B., Keller K., Cruz R., Danson M.J., Hough D.W.,  
 RA Maddock D.G., Jablonksi P.E., Krebs M.P., Angvine C.M., Dale H.,  
 RA Iendobarger T.A., Beck R.F., Penlechner M., Spudich J.L., Jung K.-H.,  
 RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,  
 RA Ehardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.;  
 RT "Genome sequence of Halobacterium species NRC-1."  
 RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).  
 DR EMBL; AF005071; AAG19867.1; -.  
 KW Complete proteome.  
 SQ SEQUENCE 473 AA; 50486 MW; 665C7A7C655E22A7 CRC64;

Query Match 86.5%; Score 32; DB 17; Length 473;  
 Best Local Similarity 85.7%; Pred. No. 86;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GETREPL 7  
 Db 14 GETREPL 20

RESULT 9  
 ID 046138 PRELIMINARY; PRT; 602 AA.  
 AC 046138;  
 DT 01-JUN-1998 (TREMBLrel. 06, Created)  
 DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE Intermediate filament protein.  
 OS Lineus sanguineus (Ribdon worm).  
 OC Eukaryota; Metazoa; Nemertea; Anopla; Heteronemertea; Lineidae;  
 OC Lineus  
 CX NCBI\_TaxID=48190;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Boyenheult M.;  
 RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.  
 DR EMBL; X99996; CA68255.1; -.  
 DR InterPro; IPR000977; DNA\_1ligase.  
 DR InterPro; IPR001664; IF.  
 DR InterPro; IPR001322; IF\_tail.

DR Pfam; PF00038; filament; 1.  
 DR Pfam; PF00932; IF\_tail; 1.  
 DR PROSITE; PS00697; DNA\_LIGASE\_A1; UNKNOWN\_1.  
 DR PROSITE; PS00226; IF\_1.  
 KM Coiled coil; Intermediate filament.  
 SQ SEQUENCE 602 AA; 67671 MW; 2A46BD1DD76AABE CRC64;  
 Query Match 86.5%; Score 32; DB 5; Length 602;  
 Best Local Similarity 71.4%; Pred. No. 1.1e+02;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 GETREPL 7  
 Db 31 GQREPL 37

RESULT 10  
 ID 095251 PRELIMINARY; PRT; 611 AA.  
 AC 095251;  
 DT 01-MAY-1999 (TREMBLrel. 10, Created)  
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
 DE Histone acetyltransferase (Histone acetyltransferase MYST2).  
 GN HB01 OR HBOA OR MYST2.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 CX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Iizuka M., Stillman B.;  
 RT "Histone acetyltransferase HB01 interacts with the ORC1 subunit of the  
 RT human initiator protein."  
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Jian J., Guangtao L., Guangwei D., Yan Z., Jianhe C., Jiangang Y.,  
 RA Jiao Q.;  
 RT "Cloning and Identifying Histone Acetyltransferase HBoa."  
 RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Borrow J., Housman D.E.;  
 RT "Structure and function of the human MYST family: MOZ2, MYST1 and  
 RT MYST2."  
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF074606; AAC9368.1; -.  
 DR EMBL; AF140360; AAD42348.1; -.  
 DR EMBL; AF217502; AAL56649.1; -.  
 DR InterPro; IPR002717; MOZ SAS.  
 DR InterPro; IPR000822; Znf\_C2H2.  
 DR InterPro; IPR002515; Znf\_C2HC.  
 DR Pfam; PF01853; MOZ SAS; 1.  
 DR Pfam; PF00096; zfc-C2H2; 1.  
 DR Pfam; PF01530; zfc-C2HC; 1.  
 DR SMART; SM00355; Znf\_C2H2; 1.  
 KM DNA-Binding; Transferase; Zinc-finger.  
 SQ SEQUENCE 611 AA; 70642 MW; 8368E7C4F07D8D7C CRC64;

Query Match 86.5%; Score 32; DB 4; Length 611;  
 Best Local Similarity 85.7%; Pred. No. 1.1e+02;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GETREPL 7  
 Db 291 GNTREPL 297

RESULT 11  
 Q9RW6 PRELIMINARY; PRT; 736 AA.  
 ID Q9RW6  
 AC Q9RW6;



```

DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 20, last annotation update)
DE Hypothetical protein DR0905.
GN DR0905.
OS Deinococcus radiodurans.
OC Bacteria; Thermus/Deinococcus group; Deinococci; Deinococcales;
OC Deinococcaceae; Deinococcus.
OX NCBI_TaxId=1299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RI.
RX MEDLINE=20036896; PubMed=10567266;
RA White O., Sisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwin M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Uitterback T., Zalewski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the radioreistant bacterium Deinococcus
RT radiodurans RI.";
RL Science 286:1571-1577(1999).
DR EMBL; AE001943; AAF10482.1; -.
DR TIGR; DR0905; -.
DR InterPro; IPR001254; Ser protease Try.
DR PROSITE; PS00135; TRYPSIN_SER; UNKNOWN 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 736 AA; 77444 MW; B501A9A187B928FB CRC64;

Query Match
Best Local Similarity 86.5%; Score 32; DB 16; Length 736;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GETREPL 7
DB 605 GEARREPL 611

RESULT 12
ID 09K1B4 PRELIMINARY; PRT; 796 AA.
AC 09K1B4;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, last annotation update)
DE DNA gyrase subunit B.
GN NMB0212.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxId=491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MC58 / SEROGROUP B;
RX MEDLINE=20175755; PubMed=10710307;
RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.B.,
RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA Nelson W.C., Gwin M.L., Deboy R., Peterson J.D., Hickey E.K.,
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mason T., Ciecko A., Parksey D.S., Blair E., Ciftone H., Clark E.B.,
RA Cotton M.D., Uitterback T.R., Khouri H., Qin H., Vamathevan J.,
RA Gill J., Scariato V., Maignani V., Piazza M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
RT MC58.";
RL Science 287:1809-1815(2000).
CC -1- CATALYTIC ACTIVITY: ATP-DEPENDENT BREAKAGE, PASSAGE AND REJOINING
CC OF DOUBLE-STRANDED DNA.
CC -1- SIMILARITY: BELONGS TO THE TYPE II TOPOISOMERASE FAMILY.
CC EMBL; AE002378; AAF40668.1; -.
DR HSSP; P06982; 1A76.
DR TIGR; NMB0212; -.
DR InterPro; IPR003594; ATPbind_ATPase.

```

```

DR InterPro; IPR002936; DNAPrim_toprim.
DR InterPro; IPR002288; DNA_gyraseB_C.
DR InterPro; IPR001241; DNA_topoisomII.
DR InterPro; IPR004359; His_KIN_sig.
DR Pfam; PR00204; DNA_gyraseB_I.
DR Pfam; PR00986; DNA_gyraseB_C; 1.
DR Pfam; PR02518; HATPase_C; 1.
DR Pfam; PR01751; Toprim; 1.
DR PRINTS; PR00418; TP12FAMILY.
DR ProDom; PD149633; DNA_gyraseB_C; 1.
DR SMART; SM00387; HATPase_C; 1.
DR SMART; SM00433; TOP2C; 1.
DR TIGRFAMs; TIGR01059; gybB; 1.
DR PROSITE; PS00177; TOPOISOMERASE_II; 1.
KW ATP-binding; Isomerase; Topoisomerase; Complete proteome.
SQ SEQUENCE 796 AA; 88187 MW; 0DD26B6C8807A6FD CRC64;

Query Match
Best Local Similarity 86.5%; Score 32; DB 16; Length 796;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GETREPL 7
DB 152 GETREPL 158

RESULT 13
ID 09JX55 PRELIMINARY; PRT; 796 AA.
AC 09JX55;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, last annotation update)
DE DNA gyrase subunit B (EC 5.99.1.3).
GN GYRB OR NMA0056.
OS Neisseria meningitidis (serogroup A).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxId=65699;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=22491 / SEROGROUP A / SEROTYPE 4A;
RX MEDLINE=2022556; PubMed=10761919;
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
RA Davies R.M., Davis P., Devlin K., Feltham T., Hamlin N., Holtroyd S.,
RA Tagels K., Leather S., Moutle S., Mungall K., Quail M.A.,
RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
RA Whitehead S., Spratt B.G., Barrall B.G.;
RT "Complete DNA sequence of a serogroup A strain of Neisseria
RT meningitidis 22491.";
RL Nature 404:502-506(2000).
CC -1- CATALYTIC ACTIVITY: ATP-DEPENDENT BREAKAGE, PASSAGE AND REJOINING
CC OF DOUBLE-STRANDED DNA.
CC -1- SIMILARITY: BELONGS TO THE TYPE II TOPOISOMERASE FAMILY.
CC EMBL; A1162752; CAB83373.1; -.
DR HSSP; P06982; 1A76.
DR InterPro; IPR003594; ATPbind_ATPase.
DR InterPro; IPR002936; DNAPrim_toprim.
DR InterPro; IPR002288; DNA_gyraseB_C.
DR InterPro; IPR001241; DNA_topoisomII.
DR InterPro; IPR004359; His_KIN_sig.
DR Pfam; PR00204; DNA_gyraseB_I.
DR Pfam; PR00986; DNA_gyraseB_C; 1.
DR Pfam; PR02518; HATPase_C; 1.
DR Pfam; PR01751; Toprim; 1.
DR PRINTS; PR00418; TP12FAMILY.
DR ProDom; PD149633; DNA_gyraseB_C; 1.
DR SMART; SM00387; HATPase_C; 1.
DR SMART; SM00433; TOP2C; 1.
DR TIGRFAMs; TIGR01059; gybB; 1.
DR PROSITE; PS00177; TOPOISOMERASE_II; 1.
KW ATP-binding; Isomerase; Topoisomerase; Complete proteome.
SQ SEQUENCE 796 AA; 88073 MW; 5316314712BC2F3 CRC64;

```



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2003, 13:57:26 ; Search time 8.61539 Seconds  
(without alignments)  
23.906 Million cell updates/sec

Title: US-09-807-949A-108

Perfect score: 37

Sequence: 1 GETREPL 7

Scoring table: BLOSUM62

Searched: Gapop 10.0, Gapext 0.5

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database:

Issued Patents AA:\*  
1: /cgn2\_6/prodata/1/aa/5A\_COMB.pep:\*  
2: /cgn2\_6/prodata/1/aa/5B\_COMB.pep:\*  
3: /cgn2\_6/prodata/1/aa/6A\_COMB.pep:\*  
4: /cgn2\_6/prodata/1/aa/6B\_COMB.pep:\*  
5: /cgn2\_6/prodata/1/aa/PCUS\_COMB.pep:\*  
6: /cgn2\_6/prodata/1/aa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	37	100.0	7	4	US-09-178-115-108 Sequence 108, App
2	37	100.0	7	4	US-09-177-776-108 Sequence 108, App
3	31	83.8	7	4	US-09-178-115-107 Sequence 107, App
4	31	83.8	7	4	US-09-177-776-107 Sequence 107, App
5	31	83.8	7	4	US-09-084-605B-29 Sequence 29, App
6	31	83.8	7	4	US-09-438-150-6 Sequence 6, App
7	31	83.8	444	4	US-08-822-774-52 Sequence 52, App
8	31	83.8	444	4	US-09-632-711-52 Sequence 52, App
9	31	83.8	444	4	US-09-632-703B-52 Sequence 52, App
10	31	83.8	444	4	US-09-632-702-52 Sequence 52, App
11	30	81.1	12	1	US-08-471-780C-35 Sequence 35, App
12	30	81.1	12	1	US-08-467-282B-35 Sequence 35, App
13	30	81.1	12	1	US-08-471-282A-35 Sequence 35, App
14	30	81.1	12	2	US-08-466-710C-35 Sequence 35, App
15	30	81.1	12	3	US-08-468-739C-35 Sequence 35, App
16	30	81.1	18	1	US-08-471-780C-36 Sequence 36, App
17	30	81.1	18	1	US-08-467-282B-36 Sequence 36, App
18	30	81.1	18	2	US-08-471-282A-36 Sequence 36, App
19	30	81.1	18	2	US-08-466-710C-36 Sequence 36, App
20	30	81.1	18	3	US-08-468-739C-36 Sequence 36, App
21	30	81.1	43	1	US-08-471-780C-54 Sequence 54, App
22	30	81.1	43	1	US-08-467-282B-54 Sequence 54, App
23	30	81.1	43	2	US-08-471-282A-54 Sequence 54, App
24	30	81.1	43	2	US-08-466-710C-54 Sequence 54, App
25	30	81.1	43	3	US-08-468-739C-54 Sequence 54, App
26	30	81.1	120	4	US-09-188-930-180 Sequence 180, App
27	29	78.4	7	4	US-09-178-115-109 Sequence 109, App

28	29	78.4	7	4	US-09-177-776-109 Sequence 109, App
29	29	78.4	186	4	US-09-149-476-461 Sequence 461, App
30	29	78.4	563	4	US-09-131-750-2 Sequence 2, App
31	29	78.4	722	1	US-08-158-232-51 Sequence 51, App
32	29	78.4	722	2	US-08-611-928-51 Sequence 51, App
33	29	78.4	722	3	US-09-173-891-51 Sequence 51, App
34	28	75.7	142	4	US-09-189-527-9 Sequence 9, App
35	28	75.7	271	4	US-09-189-527-11 Sequence 11, App
36	28	75.7	333	4	US-08-960-780-44 Sequence 44, App
37	28	75.7	333	4	US-09-073-898-44 Sequence 44, App
38	28	75.7	570	3	US-08-826-246-2 Sequence 2, App
39	28	75.7	570	3	US-08-944-495-2 Sequence 2, App
40	28	75.7	570	3	US-09-126-640-7 Sequence 7, App
41	28	75.7	570	4	US-08-925-588-2 Sequence 2, App
42	28	75.7	570	4	US-09-288-292A-7 Sequence 7, App
43	28	75.7	571	4	US-09-071-035-168 Sequence 168, App
44	28	75.7	593	4	US-09-071-035-166 Sequence 166, App
45	28	75.7	719	4	US-08-765-907A-15 Sequence 15, App

#### ALIGNMENTS

RESULT 1  
US-09-178-115-108  
Sequence 108, Application US/09178115  
Patent No. 6297041  
GENERAL INFORMATION:  
APPLICANT: Zavada, Jan  
APPLICANT: Pastorekova, Silvia  
APPLICANT: Pastorek, Jaromir  
TITLE OF INVENTION: MN Gene and Protein  
FILE REFERENCE: D-0021, 5A  
CURRENT APPLICATION NUMBER: US/09/178, 115  
EARLIER FILING DATE: 1998-10-23  
EARLIER APPLICATION NUMBER: 09/177, 776  
EARLIER FILING DATE: 1998-10-23  
EARLIER APPLICATION NUMBER: 08/787, 739  
EARLIER FILING DATE: 1997-01-24  
EARLIER APPLICATION NUMBER: 08/485, 049  
EARLIER FILING DATE: 1995-06-07  
EARLIER APPLICATION NUMBER: 08/486, 756  
EARLIER FILING DATE: 1995-06-07  
EARLIER APPLICATION NUMBER: 08/485, 863  
EARLIER FILING DATE: 1995-06-07  
EARLIER APPLICATION NUMBER: 08/481, 658  
EARLIER FILING DATE: 1995-06-07  
EARLIER APPLICATION NUMBER: 08/485, 862  
EARLIER FILING DATE: 1995-06-07  
EARLIER APPLICATION NUMBER: 08/485, 863  
EARLIER FILING DATE: 1995-06-07  
EARLIER APPLICATION NUMBER: 08/487, 077  
EARLIER FILING DATE: 1995-06-07  
EARLIER APPLICATION NUMBER: 08/260, 190  
EARLIER FILING DATE: 1994-06-15  
EARLIER APPLICATION NUMBER: 06/177, 093  
EARLIER FILING DATE: 1993-12-30  
EARLIER APPLICATION NUMBER: 07/964, 589  
EARLIER FILING DATE: 1992-10-21  
EARLIER APPLICATION NUMBER: PV-709-92  
EARLIER FILING DATE: 1992-03-11  
NUMBER OF SEQ ID NOS: 116  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 108  
LENGTH: 7  
TYPE: PRT  
ORGANISM: HUMAN  
US-09-178-115-108  
Query Match 100.0%; Score 37; DB 4; Length 7;  
Best local Similarity 100.0%; Pred. No. 1.9e+05;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GETREPL 7  
|||||  
Db 1 GETREPL 7

## RESULT 2

US-09-177-776-108  
; Sequence 108, Application US/09177776A  
; Patent No. 6297051  
; GENERAL INFORMATION:  
; APPLICANT: Zavada, Jan  
; APPLICANT: Pastorek, Silvia  
; APPLICANT: Pastorek, Jaromir  
; TITLE OF INVENTION: MN Gene and Protein  
; FILE REFERENCE: D-0021.5A  
; CURRENT APPLICATION NUMBER: US/09/177,776A  
; EARLIER APPLICATION NUMBER: 08/787,739  
; EARLIER FILING DATE: 1997-01-24  
; EARLIER APPLICATION NUMBER: 08/485,049  
; EARLIER FILING DATE: 1995-06-07  
; EARLIER APPLICATION NUMBER: 08/486,756  
; EARLIER FILING DATE: 1995-06-07  
; EARLIER APPLICATION NUMBER: 08/477,504  
; EARLIER FILING DATE: 1995-06-07  
; EARLIER APPLICATION NUMBER: 08/481,658  
; EARLIER FILING DATE: 1995-06-07  
; EARLIER APPLICATION NUMBER: 08/485,862  
; EARLIER FILING DATE: 1995-06-07  
; EARLIER APPLICATION NUMBER: 08/485,863  
; EARLIER FILING DATE: 1995-06-07  
; EARLIER APPLICATION NUMBER: 08/487,077  
; EARLIER FILING DATE: 1995-06-07  
; EARLIER APPLICATION NUMBER: 08/260,190  
; EARLIER FILING DATE: 1994-06-15  
; EARLIER APPLICATION NUMBER: 08/177,093  
; EARLIER FILING DATE: 1993-12-30  
; EARLIER APPLICATION NUMBER: 07/964,589  
; EARLIER FILING DATE: 1992-10-21  
; EARLIER APPLICATION NUMBER: PV-709-92  
; EARLIER FILING DATE: 1992-03-11  
; NUMBER OF SEQ ID NOS: 116  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 108  
; LENGTH: 7  
; TYPE: PRT  
; ORGANISM: HUMAN  
US-09-177-776-108

Query Match 100.0%; Score 37; DB 4; Length 7;  
Best Local Similarity 100.0%; Pred. No. 1.9e+05;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GETREPL 7  
|||||  
Db 1 GETREPL 7

## RESULT 3

US-09-178-115-107  
; Sequence 107, Application US/09178115  
; Patent No. 6297041  
; GENERAL INFORMATION:  
; APPLICANT: Zavada, Jan  
; APPLICANT: Pastorek, Silvia  
; APPLICANT: Pastorek, Jaromir  
; TITLE OF INVENTION: MN Gene and Protein  
; FILE REFERENCE: D-0021.5A  
; CURRENT APPLICATION NUMBER: US/09/178,115  
; EARLIER APPLICATION NUMBER: 09/177,776  
; EARLIER FILING DATE: 1998-10-23  
; EARLIER APPLICATION NUMBER: 08/787,739

; EARLIER FILING DATE: 1997-01-24  
; EARLIER APPLICATION NUMBER: 08/485,049  
; EARLIER FILING DATE: 1995-06-07  
; EARLIER APPLICATION NUMBER: 08/486,756  
; EARLIER FILING DATE: 1995-06-07  
; EARLIER APPLICATION NUMBER: 08/477,504  
; EARLIER FILING DATE: 1995-06-07  
; EARLIER APPLICATION NUMBER: 08/481,658  
; EARLIER FILING DATE: 1995-06-07  
; EARLIER APPLICATION NUMBER: 08/485,862  
; EARLIER FILING DATE: 1995-06-07  
; EARLIER APPLICATION NUMBER: 08/485,863  
; EARLIER FILING DATE: 1995-06-07  
; EARLIER APPLICATION NUMBER: 08/487,077  
; EARLIER FILING DATE: 1995-06-07  
; EARLIER APPLICATION NUMBER: 08/260,190  
; EARLIER FILING DATE: 1994-06-15  
; EARLIER APPLICATION NUMBER: 08/177,093  
; EARLIER FILING DATE: 1993-12-30  
; EARLIER APPLICATION NUMBER: 07/964,589  
; EARLIER FILING DATE: 1992-10-21  
; EARLIER APPLICATION NUMBER: PV-709-92  
; EARLIER FILING DATE: 1992-03-11  
; NUMBER OF SEQ ID NOS: 116  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 107  
; LENGTH: 7  
; TYPE: PRT  
; ORGANISM: HUMAN  
US-09-178-115-107

Query Match 83.8%; Score 31; DB 4; Length 7;  
Best Local Similarity 85.7%; Pred. No. 1.9e+05;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GETREPL 7  
|||||  
Db 1 GETREPL 7

## RESULT 4

US-09-177-776-107  
; Sequence 107, Application US/09177776A  
; Patent No. 6297051  
; GENERAL INFORMATION:  
; APPLICANT: Zavada, Jan  
; APPLICANT: Pastorek, Silvia  
; APPLICANT: Pastorek, Jaromir  
; TITLE OF INVENTION: MN Gene and Protein  
; FILE REFERENCE: D-0021.5A  
; CURRENT APPLICATION NUMBER: US/09/177,776A  
; EARLIER APPLICATION NUMBER: 08/787,739  
; EARLIER FILING DATE: 1997-01-24  
; EARLIER APPLICATION NUMBER: 08/485,049  
; EARLIER FILING DATE: 1995-06-07  
; EARLIER APPLICATION NUMBER: 08/486,756  
; EARLIER FILING DATE: 1995-06-07  
; EARLIER APPLICATION NUMBER: 08/477,504  
; EARLIER FILING DATE: 1995-06-07  
; EARLIER APPLICATION NUMBER: 08/481,658  
; EARLIER FILING DATE: 1995-06-07  
; EARLIER APPLICATION NUMBER: 08/485,862  
; EARLIER FILING DATE: 1995-06-07  
; EARLIER APPLICATION NUMBER: 08/485,863  
; EARLIER FILING DATE: 1995-06-07  
; EARLIER APPLICATION NUMBER: 08/487,077  
; EARLIER FILING DATE: 1995-06-07  
; EARLIER APPLICATION NUMBER: 08/260,190  
; EARLIER FILING DATE: 1994-06-15  
; EARLIER APPLICATION NUMBER: 08/177,093  
; EARLIER FILING DATE: 1993-12-30  
; EARLIER APPLICATION NUMBER: 07/964,589

EARLIER FILING DATE: 1992-10-21  
EARLIER APPLICATION NUMBER: PV-709-92  
EARLIER FILING DATE: 1992-03-11  
NUMBER OF SEQ ID NOS: 116  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 107  
LENGTH: 7  
TYPE: PRT  
ORGANISM: HUMAN  
US-09-177-776-107

Query Match 83.8%; Score 31; DB 4; Length 7;  
Best Local Similarity 85.7%; Pred. No. 1.9e+05;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GETREPL 7  
DB 1 GETRAPL 7

RESULT 5  
US-09-084-605B-29  
Sequence 29; Application US/09084605B  
Patent No. 6329501  
GENERAL INFORMATION:  
APPLICANT: Smith, Bruce F.  
APPLICANT: Samoilova, Tatiana  
TITLE OF INVENTION: Methods and Compositions for Targeting  
FILE REFERENCE: 5721-8  
CURRENT APPLICATION NUMBER: US/09/084,605B  
CURRENT FILING DATE: 1998-05-26  
NUMBER OF SEQ ID NOS: 29  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 29  
LENGTH: 7  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Phage display library peptides  
US-09-084-605B-29

Query Match 83.8%; Score 31; DB 4; Length 7;  
Best Local Similarity 85.7%; Pred. No. 1.9e+05;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GETREPL 7  
DB 1 GETRAPL 7

RESULT 6  
US-09-438-150-6  
Sequence 6; Application US/09438150  
Patent No. 6399575  
GENERAL INFORMATION:  
APPLICANT: Smith, Bruce F.  
APPLICANT: Samoilova, Tatiana I.  
TITLE OF INVENTION: Methods and Compositions for Targeting  
FILE REFERENCE: 5721-13  
CURRENT APPLICATION NUMBER: US/09/438,150  
CURRENT FILING DATE: 1999-11-10  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 6  
LENGTH: 7  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Phage display library peptides  
US-09-438-150-6

Query Match 83.8%; Score 31; DB 4; Length 7;  
Best Local Similarity 85.7%; Pred. No. 1.9e+05;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GETREPL 7  
DB 1 GETRAPL 7

RESULT 7  
US-08-822-774-52  
Sequence 52; Application US/08822774  
Patent No. 6183997  
GENERAL INFORMATION:  
APPLICANT: HOGREFE, Holly  
TITLE OF INVENTION: Polymerase Enhancing Factor (PEF)  
TITLE OF INVENTION: Extracts, PEF Protein Complexes, Isolated PEF Proteins,  
and Methods for Purifying and Identifying Same  
NUMBER OF SEQUENCES: 61  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: David J. Kulik, Evenson, McKeown, Edwards &  
ADDRESSER: Lenahan, P.L.L.C.  
STREET: 1200 G Street, N.W. Suite 700  
CITY: Washington  
STATE: D.C.  
ZIP: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/822,774  
FILING DATE: 21-MAR-1997  
CLASSIFICATION: 436  
ATTORNEY/AGENT INFORMATION:  
NAME: KULIK, David J.  
REGISTRATION NUMBER: 36,576  
REFERENCE/DOCKET NUMBER: 1486/43163  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 628-8800  
TELEFAX: (202) 628-8844  
INFORMATION FOR SEQ ID NO: 52:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 444 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULAR TYPE: protein  
HYPOTHEICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Escherichia coli  
US-08-822-774-52

Query Match 83.8%; Score 31; DB 4; Length 444;  
Best Local Similarity 85.7%; Pred. No. 63;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GETREPL 7  
DB 227 GPTREPL 233

RESULT 8  
US-09-632-711-52  
Sequence 52; Application US/09632711  
Patent No. 6333165  
GENERAL INFORMATION:  
APPLICANT: HOGREFE, Holly  
TITLE OF INVENTION: Polymerase Enhancing Factor (PEF)  
Extracts, PEF Protein Complexes, Isolated PEF Proteins,  
and Methods for Purifying and Identifying Same

NUMBER OF SEQUENCES: 61  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: David J. Kulik, Evenson, McKeown, Edwards &  
STREET: 1200 G Street, N.W. Suite 700  
CITY: Washington  
STATE: D.C.  
ZIP: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/632,711  
FILING DATE: 04-Aug-2000  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/822,774  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: KULIK, David J.  
REGISTRATION NUMBER: 36,576  
REFERENCE/DOCKET NUMBER: 1486/43163  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 628-8800  
TELEFAX: (202) 628-8844  
INFORMATION FOR SEQ ID NO: 52:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 444 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Escherichia coli  
SEQUENCE DESCRIPTION: SEQ ID NO: 52:  
US-09-632-711-52

Query Match 83.8%; Score 31; DB 4; Length 444;  
Best Local Similarity 85.7%; Pred. No. 63;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GETREPL 7  
DB 227 GPTREPL 233

RESULT 9  
US-09-632-703B-52  
Sequence 52, Application US/09632703B  
Patent No. 6379553  
GENERAL INFORMATION:  
APPLICANT: HOGREFE, Holly  
TITLE OF INVENTION: Polymerase Enhancing Factor (PEF) Extracts, PEF  
Protein Complexes, Isolated PEF Proteins, and Methods for F  
Identifying Same  
NUMBER OF SEQUENCES: 61  
CLASSIFICATION: <Unknown>  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Flinnegan, Henderson, Farabow, Garrett, & Dunner, L.L.P.  
STREET: 1300 I Street, N.W.  
CITY: Washington  
STATE: D.C.  
ZIP: 20005-3315  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: <Unknown>  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/632,703B  
FILING DATE: 24-Aug-2000

CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/822,774  
FILING DATE: 21-MAR-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: BARKER, M. Paul  
REGISTRATION NUMBER: 32,013  
REFERENCE/DOCKET NUMBER: 4121,0116-04  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (650) 849-6613  
TELEFAX: (650) 849-6666  
INFORMATION FOR SEQ ID NO: 52:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 444 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Escherichia coli  
SEQUENCE DESCRIPTION: SEQ ID NO: 52:  
US-09-632-703B-52

Query Match 83.8%; Score 31; DB 4; Length 444;  
Best Local Similarity 85.7%; Pred. No. 63;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GETREPL 7  
DB 227 GPTREPL 233

RESULT 10  
US-09-632-702-52  
Sequence 52, Application US/09632702  
Patent No. 6444428  
GENERAL INFORMATION:  
APPLICANT: HOGREFE, Holly  
TITLE OF INVENTION: Polymerase Enhancing Factor (PEF)  
Extracts, PEF Protein Complexes, Isolated PEF Proteins,  
and Methods for Purifying and Identifying Same  
NUMBER OF SEQUENCES: 61  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: David J. Kulik, Evenson, McKeown, Edwards &  
STREET: 1200 G Street, N.W. Suite 700  
CITY: Washington  
STATE: D.C.  
ZIP: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/632,702  
FILING DATE: 04-Aug-2000  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/822,774  
FILING DATE: 21-MAR-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: KULIK, David J.  
REGISTRATION NUMBER: 36,576  
REFERENCE/DOCKET NUMBER: 1486/43163  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 628-8800  
TELEFAX: (202) 628-8844  
INFORMATION FOR SEQ ID NO: 52:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 444 amino acids  
TYPE: amino acid

TOPOLOGY: unknown  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Escherichia coli  
SEQUENCE DESCRIPTION: SEQ ID NO: 52  
US-09-632-702-52

Query Match 83.8%; Score 31; DB 4; Length 444;  
Best Local Similarity 85.7%; Pred. No. 63;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GETREP 7  
Db 227 GQTRPL 233

RESULT 11  
US-08-471-780C-35  
Sequence 35, Application US/08471780C  
Patent No. 5759808  
GENERAL INFORMATION:  
APPLICANT: Casterman, Cecile  
TITLE OF INVENTION: Immunoglobulins Devoid of Light Chains  
NUMBER OF SEQUENCES: 130  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner  
STREET: 1300 I Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20005-3315  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/471,780C  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 530  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 08/106,944  
FILING DATE: 17-AUG-1993  
APPLICATION NUMBER: FR 92402326.0  
FILING DATE: 21-AUG-1992  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: FR 93401310.3  
FILING DATE: 21-MAY-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Potter, Jane E.R.  
REGISTRATION NUMBER: 33,332  
REFERENCE/DOCKET NUMBER: 04958.0008-00000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-408-4000  
TELEFAX: 202-408-4400  
INFORMATION FOR SEQ ID NO: 35:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 12 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FEATURE:  
NAME/KEY: Domain  
LOCATION: 1..12  
OTHER INFORMATION: /label= CH3  
US-08-471-780C-35  
Query Match 81.1%; Score 30; DB 1; Length 12;  
Best Local Similarity 83.3%; Pred. No. 2.3;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GETREP 6  
Db 1 GQTRPL 6

RESULT 12  
US-08-467-282B-35  
Sequence 35, Application US/08467282B  
Patent No. 580988  
GENERAL INFORMATION:  
APPLICANT: Casterman, Cecile  
TITLE OF INVENTION: Immunoglobulins Devoid of Light Chains  
NUMBER OF SEQUENCES: 130  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner  
STREET: 1300 I Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20005-3315  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/467,282B  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 536  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 08/106,944  
FILING DATE: 17-AUG-1993  
APPLICATION NUMBER: FR 92402326.0  
FILING DATE: 21-AUG-1992  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: FR 93401310.3  
FILING DATE: 21-MAY-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Potter, Jane E.R.  
REGISTRATION NUMBER: 33,332  
REFERENCE/DOCKET NUMBER: 04958.0008-00000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-408-4000  
TELEFAX: 202-408-4400  
INFORMATION FOR SEQ ID NO: 35:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 12 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FEATURE:  
NAME/KEY: Domain  
LOCATION: 1..12  
OTHER INFORMATION: /label= CH3  
US-08-467-282B-35  
Query Match 81.1%; Score 30; DB 1; Length 12;  
Best Local Similarity 83.3%; Pred. No. 2.3;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GETREP 6  
Db 1 GQTRPL 6  
RESULT 13  
US-08-471-282A-35  
Sequence 35, Application US/08471282A  
Patent No. 5840853

```

? GENERAL INFORMATION:
? APPLICANT: Casterman, Cecile
? APPLICANT: Hamers, Raymond
? TITLE OF INVENTION: Immunoglobulins Devoid of Light Chains
? NUMBER OF SEQUENCES: 130
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner
? STREET: 1300 I Street, N.W.
? CITY: Washington
? STATE: D.C.
? COUNTRY: USA
? ZIP: 20005-3315
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patentin Release #1.0, Version #1.25
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/471,282A
? FILING DATE: 06-JUN-1995
? CLASSIFICATION: 536
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/106,944
? FILING DATE: 17-AUG-1993
? APPLICATION NUMBER: FR 92402326.0
? FILING DATE: 21-AUG-1992
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: FR 93401310.3
? FILING DATE: 21-MAY-1993
? ATTORNEY/AGENT INFORMATION:
? NAME: Potler, Jane E.R.
? REGISTRATION NUMBER: 33,332
? REFERENCE/DOCKET NUMBER: 04958.0008-00000
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 202-408-4000
? TELEFAX: 202-408-4400
? INFORMATION FOR SEQ ID NO: 35:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 12 amino acids
? TYPE: amino acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: peptide
? FEATURE:
? NAME/KEY: Domain
? LOCATION: 1..12
? OTHER INFORMATION: /label= CH3
?
? US-08-471-282A-35
?
Query Match 81.1%; Score 30; DB 2; Length 12;
Best Local Similarity 83.3%; Pred. No. 2.3;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 GETREP 6
|:||||
Db 1 GQTRP 6

RESULT 14
US-08-466-710C-35
? Sequence 35, Application US/08466710C
? Patent No. 5674541
? GENERAL INFORMATION:
? APPLICANT: Casterman, Cecile
? APPLICANT: Hamers, Raymond
? TITLE OF INVENTION: Immunoglobulins Devoid of Light Chains
? NUMBER OF SEQUENCES: 130
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner
? STREET: 1300 I Street, N.W.
? CITY: Washington
? STATE: D.C.
? COUNTRY: USA
? ZIP: 20005-3315
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patentin Release #1.0, Version #1.25
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/466,739C
? FILING DATE: 06-JUN-1995
? CLASSIFICATION: 530
? PRIOR APPLICATION DATA:
```

```

? ZIP: 20005-3315
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patentin Release #1.0, Version #1.25
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/466,710C
? FILING DATE:
? CLASSIFICATION:
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US/08/106,944
? FILING DATE: 17-AUG-1993
? APPLICATION NUMBER: FR 92402326.0
? FILING DATE: 21-AUG-1992
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: FR 93401310.3
? FILING DATE: 21-MAY-1993
? ATTORNEY/AGENT INFORMATION:
? NAME: Potler, Jane E.R.
? REGISTRATION NUMBER: 33,332
? REFERENCE/DOCKET NUMBER: 04958.0008-00000
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 202-408-4400
? TELEFAX: 202-408-4000
? INFORMATION FOR SEQ ID NO: 35:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 12 amino acids
? TYPE: amino acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: peptide
? FEATURE:
? NAME/KEY: Domain
? LOCATION: 1..12
? OTHER INFORMATION: /label= CH3
?
? US-08-466-710C-35
?
Query Match 81.1%; Score 30; DB 2; Length 12;
Best Local Similarity 83.3%; Pred. No. 2.3;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 GETREP 6
|:||||
Db 1 GQTRP 6

RESULT 15
US-08-468-739C-35
? Sequence 35, Application US/08468739C
? Patent No. 6015695
? GENERAL INFORMATION:
? APPLICANT: Casterman, Cecile
? APPLICANT: Hamers, Raymond
? TITLE OF INVENTION: Immunoglobulins Devoid of Light Chains
? NUMBER OF SEQUENCES: 130
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner
? STREET: 1300 I Street, N.W.
? CITY: Washington
? STATE: D.C.
? COUNTRY: USA
? ZIP: 20005-3315
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patentin Release #1.0, Version #1.25
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/468,739C
? FILING DATE: 06-JUN-1995
? CLASSIFICATION: 530
? PRIOR APPLICATION DATA:
```



APPLICATION NUMBER: US 08/106,944  
FILING DATE: 17-AUG-1993  
APPLICATION NUMBER: FR 92402326.0  
FILING DATE: 21-AUG-1992  
PRIOR APPLICATION DATA: FR 93401310.3  
APPLICATION NUMBER: FR 93401310.3  
FILING DATE: 21-MAY-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Potter, Jane E. R.  
REGISTRATION NUMBER: 33,332  
REFERENCE/DOCKET NUMBER: 04958.0008-00000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-408-4000  
TELEFAX: 202-408-4400  
INFORMATION FOR SEQ ID NO: 35:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 12 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FEATURE:  
NAME/KEY: Domain  
LOCATION: 1..12  
OTHER INFORMATION: /label= CH3  
US-08-468-739C-35

Query Match 81.1%; Score 30; DB 3; Length 12;  
Best Local Similarity 83.3%; Pred. No. 2.3;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
OY 1 GETREP 6  
|:|||||  
Db 1 GOTREP 6

Search completed: January 29, 2003, 14:04:18  
Job time : 9.61539 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2003, 14:00:01 ; Search time 5.38462 Seconds  
(without alignments)  
26.232 Million cell updates/sec

Title: US-09-807-949a-108  
Perfect score: 37  
Sequence: 1 GETREPL 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 122226 seqs, 20178551 residues

Total number of hits satisfying chosen parameters: 122226

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:\*  
1: /cgn2\_6/prodata/1/pubppa/US08\_NEW\_PUB.pep:\*  
2: /cgn2\_6/prodata/1/pubppa/PC7\_NEW\_PUB.pep:\*  
3: /cgn2\_6/prodata/1/pubppa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/prodata/1/pubppa/US07\_NEW\_PUB.pep:\*  
5: /cgn2\_6/prodata/1/pubppa/US10\_NEW\_PUB.pep:\*  
6: /cgn2\_6/prodata/1/pubppa/PC7US\_PUBCOMB.pep:\*  
7: /cgn2\_6/prodata/1/pubppa/US09\_PUBCOMB.pep:\*  
8: /cgn2\_6/prodata/1/pubppa/US09\_PUBCOMB.pep:\*  
9: /cgn2\_6/prodata/1/pubppa/US09\_PUBCOMB.pep:\*  
10: /cgn2\_6/prodata/1/pubppa/US09\_PUBCOMB.pep:\*  
11: /cgn2\_6/prodata/1/pubppa/US10\_NEW\_PUB.pep:\*  
12: /cgn2\_6/prodata/1/pubppa/US10\_PUBCOMB.pep:\*  
13: /cgn2\_6/prodata/1/pubppa/US60\_NEW\_PUB.pep:\*  
14: /cgn2\_6/prodata/1/pubppa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	33	89.2	243	10 US-09-864-761-49091	Sequence 49091, A
2	31	83.8	7	10 US-09-847-137-29	Sequence 29, App
3	31	83.8	390	9 US-09-712-363-192	Sequence 192, App
4	30	81.1	231	9 US-09-252-150-16	Sequence 16, App
5	30	81.1	231	9 US-09-252-150-17	Sequence 17, App
6	30	81.1	234	9 US-09-252-150-21	Sequence 21, App
7	30	81.1	246	9 US-09-252-150-18	Sequence 18, App
8	30	81.1	248	9 US-09-252-150-19	Sequence 19, App
9	30	81.1	250	9 US-09-252-150-20	Sequence 20, App
10	30	81.1	271	10 US-09-925-297-545	Sequence 545, App
11	30	81.1	364	9 US-09-764-868-1087	Sequence 1087, App
12	30	81.1	408	10 US-09-057-951-4	Sequence 4, App
13	30	81.1	408	10 US-10-105-150-4	Sequence 4, App
14	30	81.1	430	10 US-09-057-951-2	Sequence 2, App
15	30	81.1	430	10 US-09-836-607-2	Sequence 2, App
16	30	81.1	430	10 US-10-105-150-2	Sequence 2, App
17	30	81.1	612	9 US-09-764-868-663	Sequence 663, App
18	30	81.1	820	9 US-09-989-442-118	Sequence 118, App
19	29	78.4	327	10 US-09-815-242-5720	Sequence 5720, App

20	29	78.4	333	10 US-09-815-242-12481	Sequence 12481, A
21	29	78.4	458	10 US-09-739-457-4	Sequence 4, App
22	29	78.4	490	10 US-09-739-457-3	Sequence 3, App
23	29	78.4	490	10 US-09-739-457-6	Sequence 6, App
24	29	78.4	490	10 US-09-739-457-7	Sequence 7, App
25	29	78.4	722	10 US-09-833-533A-10	Sequence 10, App
26	28	75.7	128	10 US-09-864-761-35529	Sequence 35529, A
27	28	75.7	208	10 US-09-864-761-35865	Sequence 35865, A
28	28	75.7	333	10 US-09-850-351A-44	Sequence 44, App
29	28	75.7	375	9 US-09-738-626-4744	Sequence 4744, App
30	28	75.7	416	10 US-09-815-242-11301	Sequence 11301, A
31	28	75.7	416	10 US-09-881-752A-172	Sequence 172, App
32	28	75.7	565	9 US-09-766-954A-2	Sequence 2, App
33	28	75.7	570	8 US-08-825-486-2	Sequence 2, App
34	28	75.7	570	8 US-08-870-434-7	Sequence 7, App
35	28	75.7	570	10 US-09-372-044-2	Sequence 2, App
36	28	75.7	681	10 US-09-838-539-8	Sequence 8, App
37	28	75.7	1080	10 US-09-900-237-30	Sequence 30, App
38	28	75.7	2186	10 US-09-927-668-2	Sequence 2, App
39	28	75.7	2472	10 US-09-815-242-5064	Sequence 5064, App
40	27	73.0	8	10 US-09-243-079-50	Sequence 50, App
41	27	73.0	56	10 US-09-864-761-43359	Sequence 43359, A
42	27	73.0	136	10 US-09-925-299-1246	Sequence 1246, App
43	27	73.0	166	9 US-09-764-868-946	Sequence 946, App
44	27	73.0	175	9 US-09-738-626-4757	Sequence 4757, App
45	27	73.0	198	9 US-09-981-353-138	Sequence 138, App

#### ALIGNMENTS

RESULT 1  
US-09-864-761-49091  
Sequence 49091, Application US/09864761  
Patent No. US2002046763A1  
GENERAL INFORMATION:  
APPLICANT: Penn, Sharon G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
APPLICANT: Chen, Wenhang  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
FILE REFERENCE: Aecmca-X-1  
CURRENT APPLICATION NUMBER: US/09/864,761  
CURRENT FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: US 60/180,312  
PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: US 60/207,456  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: US 09/632,366  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: GB 24263.6  
PRIOR FILING DATE: 2000-10-04  
PRIOR APPLICATION NUMBER: US 60/236,359  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: PCT/US01/00666  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00667  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00664  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00669  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00665  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00668  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00663  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00662  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00661  
PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00670  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: US 09/608,408  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774,203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 49091  
LENGTH: 243  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AC005678.1  
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.3  
OTHER INFORMATION: EST\_HUMAN HIT: AUI41046.1, EVALU0 1.00e-66  
OTHER INFORMATION: SWISSPROT HIT: O43309, EVALU0 0.00e+00  
US-09-864-761-49091

Query Match 89.2%; Score 31; DB 10; Length 243;  
Best Local Similarity 100.0%; Pred. No. 12;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GETREP 6  
Db 41 GETREP 46

RESULT 2  
US-09-947-137-29  
Sequence 29, Application US/09947137  
Patent No. US20020137023A1  
GENERAL INFORMATION:  
APPLICANT: Smith, Bruce F.  
APPLICANT: Samoilova, Tatiana  
TITLE OF INVENTION: Methods and Compositions for Targeting  
FILE REFERENCE: 5721-8  
CURRENT APPLICATION NUMBER: US/09/947,137  
CURRENT FILING DATE: 2001-09-05  
PRIOR APPLICATION NUMBER: 09/084,605  
PRIOR FILING DATE: 1998-05-26  
NUMBER OF SEQ ID NOS: 29  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 29  
LENGTH: 7  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Phage display library peptides  
US-09-947-137-29

Query Match 83.8%; Score 31; DB 10; Length 7;  
Best Local Similarity 85.7%; Pred. No. 1e+05;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GETREP 7  
Db 1 GETREP 7

RESULT 3  
US-09-712-363-192  
Sequence 192, Application US/09712363  
Patent No. US20020164588A1  
GENERAL INFORMATION:  
APPLICANT: Eisenberg, David  
APPLICANT: Rotstein, Sergio H.  
APPLICANT: Marcotte, Edward M.  
TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND  
INTERACTIONS OF PROTEINS BY COMPARATIVE ANALYSIS

FILE REFERENCE: 07419-032001  
CURRENT APPLICATION NUMBER: US/09/712,363  
CURRENT FILING DATE: 2000-11-13  
PRIOR APPLICATION NUMBER: PCT/US00/02246  
PRIOR FILING DATE: 2000-01-28  
PRIOR APPLICATION NUMBER: 60/179,531  
PRIOR FILING DATE: 2000-02-01  
PRIOR APPLICATION NUMBER: 60/117,844  
PRIOR FILING DATE: 1999-01-29  
PRIOR APPLICATION NUMBER: 60/118,206  
PRIOR FILING DATE: 1999-02-01  
PRIOR APPLICATION NUMBER: 60/126,593  
PRIOR FILING DATE: 1999-03-26  
PRIOR APPLICATION NUMBER: 60/134,093  
PRIOR FILING DATE: 1999-05-14  
PRIOR APPLICATION NUMBER: 60/134,092  
PRIOR FILING DATE: 1999-05-14  
PRIOR APPLICATION NUMBER: 60/165,124  
PRIOR FILING DATE: 1999-11-12  
PRIOR APPLICATION NUMBER: 60/165,086  
PRIOR FILING DATE: 1999-11-12  
NUMBER OF SEQ ID NOS: 292  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 192  
LENGTH: 390  
TYPE: PRT  
ORGANISM: Mycobacterium tuberculosis  
US-09-712-363-192

Query Match 83.8%; Score 31; DB 9; Length 390;  
Best Local Similarity 71.4%; Pred. No. 50;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 GETREP 7  
Db 110 GETREP 116

RESULT 4  
US-09-252-150-16  
Sequence 16, Application US/09252150A  
Patent No. US20020155604A1  
GENERAL INFORMATION:  
APPLICANT: Ledbetter, Jeffrey A.  
APPLICANT: Hayden Ledbetter, Martha  
APPLICANT: Brady, William A.  
APPLICANT: Grosmaire, Laura S.  
APPLICANT: Law, Che-Leung  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR REGULATING  
LYMPHOCYTE ACTIVATION  
FILE REFERENCE: 9113-0019-999  
CURRENT APPLICATION NUMBER: US/09/252,150A  
CURRENT FILING DATE: 1999-02-18  
EARLIER APPLICATION NUMBER: US 60/075,274  
EARLIER FILING DATE: 1998-02-19  
EARLIER APPLICATION NUMBER: US 60/108,683  
EARLIER FILING DATE: 1998-11-16  
NUMBER OF SEQ ID NOS: 80  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 16  
LENGTH: 231  
TYPE: PRT  
ORGANISM: llama llama  
US-09-252-150-16

Query Match 81.1%; Score 30; DB 9; Length 231;  
Best Local Similarity 83.3%; Pred. No. 46;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GETREP 6  
Db 123 GETREP 128

```
RESULT 5
US-09-252-150-17
; Sequence 17, Application US/09252150A
; Patent No. US20020155604A1
; GENERAL INFORMATION:
; APPLICANT: Ledbetter, Jeffrey A.
; APPLICANT: Hayden Ledbetter, Martha
; APPLICANT: Brady, William A.
; APPLICANT: Grosmaire, Laura S.
; APPLICANT: Law, Che-Leung
; APPLICANT: Dua, Raj
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR REGULATING
; FILE OF INVENTION: LYMPHOCYTE ACTIVATION
; FILE REFERENCE: 9113-0019-999
; CURRENT APPLICATION NUMBER: US/09/252,150A
; EARLIER FILING DATE: 1999-02-18
; EARLIER APPLICATION NUMBER: US 60/075,274
; EARLIER FILING DATE: 1998-02-19
; EARLIER APPLICATION NUMBER: US 60/108,683
; EARLIER FILING DATE: 1998-11-16
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 17
; LENGTH: 231
; TYPE: PRT
; ORGANISM: Llama llama
US-09-252-150-17
```

```
Query Match
Best Local Similarity 81.1%; Score 30; DB 9; Length 231;
Best Local Similarity 83.3%; Pred. No. 46;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

QY 1 GETREP 6  
|:|||||  
Db 123 GOTREP 128

```
RESULT 6
US-09-252-150-21
; Sequence 21, Application US/09252150A
; Patent No. US20020155604A1
; GENERAL INFORMATION:
; APPLICANT: Ledbetter, Jeffrey A.
; APPLICANT: Hayden Ledbetter, Martha
; APPLICANT: Brady, William A.
; APPLICANT: Grosmaire, Laura S.
; APPLICANT: Law, Che-Leung
; APPLICANT: Dua, Raj
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR REGULATING
; FILE OF INVENTION: LYMPHOCYTE ACTIVATION
; FILE REFERENCE: 9113-0019-999
; CURRENT APPLICATION NUMBER: US/09/252,150A
; EARLIER FILING DATE: 1999-02-18
; EARLIER APPLICATION NUMBER: US 60/075,274
; EARLIER FILING DATE: 1998-02-19
; EARLIER APPLICATION NUMBER: US 60/108,683
; EARLIER FILING DATE: 1998-11-16
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 21
; LENGTH: 214
; TYPE: PRT
; ORGANISM: Llama llama
US-09-252-150-21
```

```
Query Match
Best Local Similarity 81.1%; Score 30; DB 9; Length 234;
Best Local Similarity 83.3%; Pred. No. 46;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

QY 1 GETREP 6  
|:|||||

Db 126 GOTREP 131

```
RESULT 7
US-09-252-150-18
; Sequence 18, Application US/09252150A
; Patent No. US20020155604A1
; GENERAL INFORMATION:
; APPLICANT: Ledbetter, Jeffrey A.
; APPLICANT: Hayden Ledbetter, Martha
; APPLICANT: Brady, William A.
; APPLICANT: Grosmaire, Laura S.
; APPLICANT: Law, Che-Leung
; APPLICANT: Dua, Raj
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR REGULATING
; FILE OF INVENTION: LYMPHOCYTE ACTIVATION
; FILE REFERENCE: 9113-0019-999
; CURRENT APPLICATION NUMBER: US/09/252,150A
; EARLIER FILING DATE: 1999-02-18
; EARLIER APPLICATION NUMBER: US 60/075,274
; EARLIER FILING DATE: 1998-02-19
; EARLIER APPLICATION NUMBER: US 60/108,683
; EARLIER FILING DATE: 1998-11-16
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 18
; LENGTH: 246
; TYPE: PRT
; ORGANISM: Llama llama
US-09-252-150-18
```

```
Query Match
Best Local Similarity 81.1%; Score 30; DB 9; Length 246;
Best Local Similarity 83.3%; Pred. No. 49;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

QY 1 GETREP 6  
|:|||||  
Db 138 GOTREP 143

```
RESULT 8
US-09-252-150-19
; Sequence 19, Application US/09252150A
; Patent No. US20020155604A1
; GENERAL INFORMATION:
; APPLICANT: Ledbetter, Jeffrey A.
; APPLICANT: Hayden Ledbetter, Martha
; APPLICANT: Brady, William A.
; APPLICANT: Grosmaire, Laura S.
; APPLICANT: Law, Che-Leung
; APPLICANT: Dua, Raj
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR REGULATING
; FILE OF INVENTION: LYMPHOCYTE ACTIVATION
; FILE REFERENCE: 9113-0019-999
; CURRENT APPLICATION NUMBER: US/09/252,150A
; EARLIER FILING DATE: 1999-02-18
; EARLIER APPLICATION NUMBER: US 60/075,274
; EARLIER FILING DATE: 1998-02-19
; EARLIER APPLICATION NUMBER: US 60/108,683
; EARLIER FILING DATE: 1998-11-16
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 19
; LENGTH: 248
; TYPE: PRT
; ORGANISM: Llama llama
US-09-252-150-19
```

```
Query Match
Best Local Similarity 81.1%; Score 30; DB 9; Length 248;
Best Local Similarity 83.3%; Pred. No. 49;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

QY 1 GETREP 6  
|:|||||

Db 140 GOTREP 145

RESULT 9  
US-09-252-150-20  
; Sequence 20, Application US/09252150A  
; Patent No. US20020155604A1  
; GENERAL INFORMATION:  
; APPLICANT: Ledbetter, Jeffrey A.  
; APPLICANT: Hayden Ledbetter, Martha  
; APPLICANT: Brady, William A.  
; APPLICANT: Grosmaire, Laura S.  
; APPLICANT: Law, Che-Leung  
; APPLICANT: Dua, Raj  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR REGULATING  
; FILE REFERENCE: 9113-0019-999  
; CURRENT APPLICATION NUMBER: US/09/252,150A  
; CURRENT FILING DATE: 1999-02-18  
; EARLIER APPLICATION NUMBER: US 60/075,274  
; EARLIER FILING DATE: 1998-02-19  
; EARLIER APPLICATION NUMBER: US 60/108,683  
; EARLIER FILING DATE: 1998-11-16  
; NUMBER OF SEQ ID NOS: 80  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 20  
; LENGTH: 250  
; TYPE: PRT  
; ORGANISM: llama llama  
US-09-252-150-20

Query Match 81.1%; Score 30; DB 9; Length 250;  
Best Local Similarity 83.3%; Pred. No. 50;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GETREP 6  
Db 142 GOTREP 147

RESULT 10  
US-09-925-297-545  
; Sequence 545, Application US/09925297  
; Patent No. US20020081659A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
; FILE REFERENCE: PA105  
; CURRENT APPLICATION NUMBER: US/09/925,297  
; CURRENT FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: PCT/US00/05989  
; PRIOR FILING DATE: 2000-03-08  
; PRIOR APPLICATION NUMBER: 60/124,270  
; PRIOR FILING DATE: 1999-03-12  
; NUMBER OF SEQ ID NOS: 928  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 545  
; LENGTH: 271  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-925-297-545

Query Match 81.1%; Score 30; DB 10; Length 271;  
Best Local Similarity 85.7%; Pred. No. 54;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GETREP 7  
Db 79 GETRVP 85

RESULT 11

US-09-764-868-1087  
; Sequence 1087, Application US/09764868  
; Patent No. US20020168711A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PT232  
; CURRENT APPLICATION NUMBER: US/09/764,868  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - refer to PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 1510  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 1087  
; LENGTH: 364  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (25)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (81)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-764-868-1087

Query Match 81.1%; Score 30; DB 9; Length 364;  
Best Local Similarity 85.7%; Pred. No. 74;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 GETREP 7  
Db 172 GETRVP 178

RESULT 12  
US-09-057-951-4  
; Sequence 4, Application US/09057951  
; Patent No. US20020025551A1  
; GENERAL INFORMATION:  
; APPLICANT: Holzman, Douglas  
; TITLE OF INVENTION: NOVEL MOLECULES OF THE T129-RELATED  
; TITLE OF INVENTION: PROTEIN FAMILY AND USES THEREOF  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: Windows95  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/057,951  
; FILING DATE: 09-APR-1998  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Melk1eJohn, Ph.D., Anita L.  
; REGISTRATION NUMBER: 35,283  
; REFERENCE/DOCKET NUMBER: 09404/046001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617/542-5070  
; TELEFAX: 617/542-8906  
; TELEX: 200154  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 408 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-09-057-951-4

Query Match 81.1%; Score 30; DB 10; Length 408;  
Best Local Similarity 83.3%; Pred. No. 84;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GETREP 6  
Db 120 GETROP 125

RESULT 13  
US-10-105-150-4  
Sequence 4, Application US/10105150  
Patent No. US20020119524A1

GENERAL INFORMATION:

APPLICANT: Holtzman, Douglas  
TITLE OF INVENTION: NOVEL MOLECULES OF THE T129-RELATED  
PROTEIN FAMILY AND USES THEREOF

NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Fish & Richardson P.C.  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02110-2804

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows95

SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/105,150  
FILING DATE: 25-Mar-2002

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/057,951  
FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:  
NAME: Melkielejohn, Ph.D., Anita L.

REGISTRATION NUMBER: 35,283  
REFERENCE/DOCKET NUMBER: 09404/046001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617/542-5070  
TELEFAX: 617/542-8906  
TELEX: 200154

INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 408 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULAR TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 4:  
US-10-105-150-4

Query Match 81.1%; Score 30; DB 12; Length 408;  
Best Local Similarity 83.3%; Pred. No. 84;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GETREP 6  
Db 120 GETROP 125

RESULT 14  
US-09-057-951-2

Sequence 2, Application US/09057951  
Patent No. US2002002551A1

GENERAL INFORMATION:

APPLICANT: Holtzman, Douglas

TITLE OF INVENTION: NOVEL MOLECULES OF THE T129-RELATED  
PROTEIN FAMILY AND USES THEREOF

NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Fish & Richardson P.C.  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02110-2804

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows95

SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/057,951  
FILING DATE: 09-APR-1998

ATTORNEY/AGENT INFORMATION:  
NAME: Melkielejohn, Ph.D., Anita L.

REGISTRATION NUMBER: 35,283  
REFERENCE/DOCKET NUMBER: 09404/046001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617/542-5070  
TELEFAX: 617/542-8906  
TELEX: 200154

INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 430 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULAR TYPE: protein  
FRAGMENT TYPE: Internal

US-09-057-951-2

Query Match 81.1%; Score 30; DB 10; Length 430;  
Best Local Similarity 83.3%; Pred. No. 89;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GETREP 6  
Db 142 GETROP 147

RESULT 15  
US-09-836-607-2

Sequence 2, Application US/09836607  
Patent No. US20020098541A1

GENERAL INFORMATION:

APPLICANT: Ni, Jian

TITLE OF INVENTION: TNFR Related Gene 12  
FILE REFERENCE: PF490P1

CURRENT APPLICATION NUMBER: US/09/836,607  
PRIOR FILING DATE: 2001-04-17

PRIOR APPLICATION NUMBER: 60/198,388  
PRIOR FILING DATE: 2000-04-19

PRIOR APPLICATION NUMBER: 09/421,112  
PRIOR FILING DATE: 1999-10-19

PRIOR APPLICATION NUMBER: 60/104,950  
PRIOR FILING DATE: 1998-10-20

NUMBER OF SEQ ID NOS: 47  
SOFTWARE: Patentin Ver. 2.0

SEQ ID NO 2

LENGTH: 430

TYPE: PRT

ORGANISM: Homo sapiens

US-09-836-607-2

Query Match 81.1%; Score 30; DB 10; Length 430;  
Best Local Similarity 83.3%; Pred. No. 89;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GETREP 6  
Db 142 GETROP 147

Thu Jan 30 11:47:38 2003

us-09-807-949a-108.rapb

Page 6

Search completed: January 29, 2003, 14:04:54  
Job time : 5.38462 secs

---



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2003, 13:51:05 ; Search time 25.3077 Seconds  
(without alignments)  
36.857 Million cell updates/sec

Title: US-09-807-949a-109  
Perfect score: 36  
Sequence: 1 GQTRSP.L 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A. Geneseq.101002.\*

1:	/SID2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2:	/SID2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3:	/SID2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4:	/SID2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5:	/SID2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6:	/SID2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7:	/SID2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8:	/SID2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9:	/SID2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10:	/SID2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11:	/SID2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12:	/SID2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13:	/SID2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14:	/SID2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15:	/SID2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16:	/SID2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17:	/SID2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18:	/SID2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19:	/SID2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20:	/SID2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21:	/SID2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22:	/SID2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23:	/SID2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	36	100.0	7	21	AA03034	MN protein CA doma
2	36	100.0	9	21	AA03058	MN protein CA doma
3	34	94.4	86	22	AA04431	Proionbacterium
4	33	91.7	179	22	AA04543	Proionbacterium
5	33	91.7	949	22	AB05631	Drosophila melanog
6	33	91.7	949	22	AB05986	Drosophila melanog
7	32	88.9	48	22	AB03116	Peptide #3767 enco
8	32	88.9	48	22	AB036319	Peptide #3825 enco
9	32	88.9	48	22	AB021674	Protein #3673 enco
10	32	88.9	48	22	AA069471	Human bone marrow

11	32	88.9	48	22	AA017314	Peptide #3748 enco
12	32	88.9	48	22	AA029812	Peptide #3849 enco
13	32	88.9	48	22	AA044993	Peptide #3675 enco
14	32	88.9	48	23	AB039102	Human peptide enco
15	32	88.9	407	22	AA097649	Human sperm protei
16	32	88.9	844	22	AB065418	Drosophila melanog
17	31	86.1	635	22	AA035616	Haemophilus influe
18	31	86.1	637	22	AA036465	Pseudomonas aerugi
19	31	86.1	665	23	AA021150	Human 52991 protei
20	31	86.1	691	23	AA022915	Human transporter
21	31	86.1	853	20	AA081079	Amino acid sequenc
22	31	86.1	866	17	AA088408	Area activator pro
23	31	86.1	866	18	AA031630	Aspergillus oryzae
24	30	83.3	7	20	AA089659	Muscle-specific pe
25	30	83.3	7	21	AA030332	MN protein CA doma
26	30	83.3	7	21	AA084992	Amino acid sequenc
27	30	83.3	7	21	AA092734	Heptapeptide 4 mim
28	30	83.3	7	21	AA053711	Amino acid sequenc
29	30	83.3	7	23	AB047722	N. meningitidis LO
30	30	83.3	83	22	AA055483	Proionbacterium
31	30	83.3	144	22	AA002295	Human polypeptide
32	30	83.3	295	23	AB038005	Staphylococcus epi
33	30	83.3	352	21	AA096261	R. eutropha toluen
34	30	83.3	352	21	AA084821	Amino acid sequenc
35	30	83.3	369	22	AA035214	Enterococcus faeca
36	30	83.3	389	22	AB077555	Thermus cytochrome
37	30	83.3	395	23	AB077556	His tagged cytochr
38	30	83.3	395	23	AB077557	His tagged cytochr
39	30	83.3	413	23	AB028144	Streptococcus poly
40	30	83.3	670	22	AB060386	Drosophila melanog
41	30	83.3	1446	22	AB029987	Novel human diagno
42	30	83.3	2146	22	AB062317	Drosophila melanog
43	30	83.3	3319	22	AB070376	Drosophila melanog
44	29	80.6	7	21	AA030333	MN protein CA doma
45	29	80.6	15	10	AA091479	Peptide D-30. Kab

## ALIGNMENTS

XX	RESULT 1
XX	ID AAB03034
XX	standard; peptide; 7 AA.
XX	AC AAB03034;
XX	25-SEP-2000 (first entry)
XX	MN protein CA domain-binding peptide, SEQ ID NO:109.
XX	MN protein; tumour associated cell adhesion molecule; oncoprotein;
KW	proteoglycan domain; PG domain; carbonic anhydrase; CA domain;
KW	abnormal expression; neoplastic disease; cancer; gene therapy;
XX	phage display library.
OS	Synthetic.
XX	WO200024913-A2.
XX	04-MAY-2000.
XX	22-OCT-1999; 99WO-US24879.
XX	23-OCT-1998; 98US-0177776.
XX	23-OCT-1998; 98US-0178115.
XX	(PARB ) BAYER CORP.
XX	(VIRB-) INST VIROLOGY.
XX	Zavada J, Pastorekova S, Pastorek J;
XX	WPI; 2000-350752/30.
XX	

PT A molecule which specifically binds to a site on MN protein  
 PT (oncoprotein) and prevents adhesion of vertebrate cells to the protein,  
 PT useful for treating preneoplastic or neoplastic diseases such as cancer

PS Claim 5, Page 71, 154pp, English.

XX The invention relates to the inhibition of cell adhesion mediated by  
 CC the MN oncoprotein (also known as the MN/CA IX isoenzyme or the MN/G250  
 CC protein). The MN protein is a tumour-associated adhesion molecule which  
 CC comprises a proteoglycan-like (PG) domain (AAB03017) which contains the  
 CC protein's binding site, and a carbonic anhydrase (CA) domain (AAB03018).  
 CC Abnormal expression of the MN protein is associated with tumorigenicity.  
 CC The invention encompasses molecules (e.g., proteins and peptides) which  
 CC which specifically bind to a site on the MN protein, thereby preventing  
 CC adhesion of vertebrate cells to the protein in a cell adhesion assay. It  
 CC also encompasses MN proteins or MN protein fragments which can be added  
 CC to the extracellular environment to prevent the adhesion of vertebrate  
 CC cells to each other. The invention also relates to the identification of  
 CC the binding site of the MN protein and to a method of identifying a site  
 CC on an MN protein to which cells adhere, comprising testing a series of  
 CC overlapping peptides from the protein in a cell adhesion assay. The  
 CC invention encompasses a vector comprising an expression control sequence  
 CC operatively linked to a nucleic acid encoding the variable domains of a  
 CC MN-specific antibody, where the domains are separated by a flexible  
 CC linker peptide (AAB03035) and the vector inhibits the growth of a  
 CC vertebrate preneoplastic or neoplastic cell that abnormally expresses MN  
 CC protein. The invention also encompasses a vector comprising a  
 CC nucleic acid encoding a cytotoxic protein or peptide operatively linked  
 CC to the MN gene promoter, which inhibits the growth of a vertebrate  
 CC preneoplastic or neoplastic cell. Also claimed is a repressor complex  
 CC that binds to the MN gene promoter (AAB52473). MN proteins and peptides,  
 CC MN-binding proteins and peptides, and expression vectors encoding such  
 CC proteins and peptides are useful for treating patients with  
 CC preneoplastic or neoplastic disease (e.g., cancers) associated with or  
 CC characterised by abnormal MN expression. Sequences AAB03032-B03034 and  
 CC AAB03055-B03058 represent synthetic phage display library peptides which  
 CC bind to the CA domain of the human MN protein (AAB03005).

XX Sequence 7 AA;

Query Match 100.0%; Score 36; DB 21; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 7.8e+05;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GQTRSP 7  
 |||||  
 Db 1 GQTRSP 7

RESULT 2

AAB03058 ID AAB03058 standard; peptide; 9 AA.

XX AAB03058;

DT 25-SEP-2000 (first entry)

XX MN protein CA domain-binding peptide, SEQ ID NO:140.

XX MN protein; tumour associated cell adhesion molecule; oncoprotein;  
 KM proteoglycan domain; PG domain; carbonic anhydrase; CA domain;  
 KM abnormal expression; neoplastic disease; cancer; gene therapy;  
 KM phage display library.

XX Synthetic.

XX Key Location/Qualifiers

FT Modified-site 9 /note= "C-terminal amide moiety"

XX WO200024913-A2.

PD 04-MAY-2000.

XX 22-OCT-1999; 99MO-US24879.

XX 23-OCT-1998; 98US-0177776.

XX 23-OCT-1998; 98US-0178115.

PA (FARB ) BAYER CORP.  
 PA (VIRO-) INST VIROLOGY.

PI Zavada J, Pastorekova S, Pastorek J;  
 XX WPI; 2000-350752/30.

PT A molecule which specifically binds to a site on MN protein  
 PT (oncoprotein) and prevents adhesion of vertebrate cells to the protein,  
 PT useful for treating preneoplastic or neoplastic diseases such as cancer

PS Example 2, Page 68, 154pp, English.

XX The invention relates to the inhibition of cell adhesion mediated by  
 CC the MN oncoprotein (also known as the MN/CA IX isoenzyme or the MN/G250  
 CC protein). The MN protein is a tumour-associated adhesion molecule which  
 CC comprises a proteoglycan-like (PG) domain (AAB03017) which contains the  
 CC protein's binding site, and a carbonic anhydrase (CA) domain (AAB03018).  
 CC Abnormal expression of the MN protein is associated with tumorigenicity.  
 CC The invention encompasses molecules (e.g., proteins and peptides) which  
 CC which specifically bind to a site on the MN protein, thereby preventing  
 CC adhesion of vertebrate cells to the protein in a cell adhesion assay. It  
 CC also encompasses MN proteins or MN protein fragments which can be added  
 CC to the extracellular environment to prevent the adhesion of vertebrate  
 CC cells to each other. The invention also relates to the identification of  
 CC the binding site of the MN protein and to a method of identifying a site  
 CC on an MN protein to which cells adhere, comprising testing a series of  
 CC overlapping peptides from the protein in a cell adhesion assay. The  
 CC invention encompasses a vector comprising an expression control sequence  
 CC operatively linked to a nucleic acid encoding the variable domains of a  
 CC MN-specific antibody, where the domains are separated by a flexible  
 CC linker peptide (AAB03035) and the vector inhibits the growth of a  
 CC vertebrate preneoplastic or neoplastic cell that abnormally expresses MN  
 CC protein. The invention also encompasses a vector comprising a  
 CC nucleic acid encoding a cytotoxic protein or peptide operatively linked  
 CC to the MN gene promoter, which inhibits the growth of a vertebrate  
 CC preneoplastic or neoplastic cell. Also claimed is a repressor complex  
 CC that binds to the MN gene promoter (AAB52473). MN proteins and peptides,  
 CC MN-binding proteins and peptides, and expression vectors encoding such  
 CC proteins and peptides are useful for treating patients with  
 CC preneoplastic or neoplastic disease (e.g., cancers) associated with or  
 CC characterised by abnormal MN expression. Sequences AAB03032-B03034 and  
 CC AAB03055-B03058 represent synthetic phage display library peptides which  
 CC bind to the CA domain of the human MN protein (AAB03005).

XX Sequence 9 AA;

Query Match 100.0%; Score 36; DB 21; Length 9;  
 Best Local Similarity 100.0%; Pred. No. 7.8e+05;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GQTRSP 7  
 |||||  
 Db 2 GQTRSP 8

RESULT 3

AAB64431 ID AAB64431 standard; Protein; 86 AA.

XX AAB64431;

DT 27-FEB-2002 (first entry)

XX Propionibacterium acnes immunogenic protein #25327.

XX	SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
KM	uveitis; endophthalmitis; bone joint; central nervous system; ELISA;
KM	inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KW	dermatological; osteopathic; neuroprotectant.
XX	
OS	Propionibacterium acnes.
PN	
XX	WO200181581-A2.
PD	
XX	01-NOV-2001.
PF	20-APR-2001; 2001WO-US12865.
PR	21-APR-2000; 2000US-199047P.
PR	02-JUN-2000; 2000US-208841P.
PR	07-JUL-2000; 2000US-216747P.
XX	
PA	(CORI-) CORIXA CORP.
XX	
PI	Sleiky YAM, Persing DH, Mitcham JL, Wang SS, Bhatia A;
PI	L'maisonneuve J, Zhang Y, Jen S, Carter D;
DR	WIJ.; 2001-616774/71.
XX	N-PDSB; AAS59643.
PT	Propionibacterium acnes polypeptides and nucleic acids useful for
PT	vaccinating against and diagnosing infections, especially useful for
PT	treating acne vulgaris -
PS	Example 1; SEQ ID No 25626; 1069pp; English.
XX	
CC	Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
CC	polypeptides. The proteins and their associated DNA sequences are used in
CC	the treatment, prevention and diagnosis of medical conditions caused by
CC	P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
CC	pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
CC	P. acnes is also involved in infections of bone, joints and the central
CC	nervous system, however it is particularly involved in the inflammatory
CC	lesions associated with acne vulgaris. A method for detecting the
CC	presence or absence of P. acnes in a patient comprises contacting a
CC	sample with a binding agent that binds to the proteins of the invention
CC	and determining the amount of bound protein in the sample. The
CC	polypeptides may be used as antigens in the production of antibodies
CC	specific for P. acnes proteins. These antibodies can be used to
CC	dysregulate expression and activity of P. acnes polypeptides and
CC	therefore treat P. acnes infections. The antibodies may also be used as
CC	diagnostic agents for determining P. acnes presence, for example, by
CC	enzyme linked immunosorbent assay (ELISA).
CC	Note: The sequence data for this patent did not form part of the printed
CC	specification, but was obtained in electronic format directly from WIPO
CC	at ftp.wipo.int/pub/published_pct_sequences.
XX	
SO	Sequence 86 AA:
OY	Query Match 94.4%; Score 34; DB 22; Length 86;
OY	Best Local Similarity 85.7%; Pred. No. 10;
OY	Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
DB	1 GQTRSP.L 7      : 65 GQTRSP.M 71
RESULT 4	
ID	AAU64543
XX	AAU64543 standard; Protein; 179 AA.
AC	AAU64543;
DT	27-FEB-2002 (first entry)
DB	Propionibacterium acnes immunogenic protein #25439.

XX	SAHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
XX	uvetis; endophthalmitis; bone; joint; central nervous system; ELISA;
KW	inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KW	dermatological; osteopathic; neuroprotectant.
XX	
OS	Propionibacterium acne.
XX	
PN	WO200181581-A2.
XX	
PD	01-NOV-2001.
XX	
PE	20-APR-2001; 2001WO-0512855.
XX	
FR	21-APR-2000; 2000US-199047P.
PR	02-JUN-2000; 2000US-208841P.
XX	07-JUL-2000; 2000US-216747P.
PA	(CORI-) CORIXA CORP.
XX	
PI	Skeiky YAM, Persing DH, Mitcham JL, Wang SS, Bhalaria A;
PI	L'maisonneuve J, Zhang Y, Jen S, Carter D;
XX	WPI; 2001-616774/71.
DR	N-PSDB; AABS59645.
XX	
PT	Propionibacterium acne polypeptides and nucleic acids useful for
PT	vaccinating against and diagnosing infections, especially useful for
PT	treating acne vulgaris -
XX	
PS	Example 1; SEQ ID NO 25738; 1069PP; English.
XX	
CC	Sequences AAU93105-AAU68017 represent Propionibacterium acne immunogenic
CC	polypeptides. The proteins and their associated DNA sequences are used in
CC	the treatment, prevention and diagnosis of medical conditions caused by
CC	P. acnes. The disorders include SAHO syndrome (synovitis, acne,
CC	pustulosis, hypertosis and osteomyelitis), uvetis and endophthalmitis.
CC	P. acnes is also involved in infections of bone, joints and the central
CC	nervous system, however it is particularly involved in the inflammatory
CC	lesions associated with acne vulgaris. A method for detecting the
CC	presence or absence of P. acnes in a patient comprises contacting a
CC	sample with a binding agent that binds to the proteins of the invention
CC	and determining the amount of bound protein in the sample. The
CC	polypeptides may be used as antigens in the production of antibodies
CC	specific for P. acnes proteins. These antibodies can be used to
CC	downregulate expression and activity of P. acnes polypeptides and
CC	therefore treat P. acnes infections. The antibodies may also be used as
CC	diagnostic agents for determining P. acnes presence, for example, by
CC	enzyme linked immunosorbent assay (ELISA).
CC	Note: The sequence data for this patent did not form part of the printed
CC	specification, but was obtained in electronic format directly from WIPO
CC	at ftp.wipo.int/pub/published_pct_sequences.
XX	
SO	Sequence 179 AA;
XX	
QY	1 GQTRSP.L 7
XX	
Db	84 GTRSP.L 90
XX	
XX	Query Match 91.7%; Score 33; DB 22; Length 179;
XX	Best Local Similarity 85.7%; Pred. No. 37;
XX	Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
XX	
XX	RESULT 5
XX	ABBS59631
ID	ABBS59631 standard; Protein; 949 AA.
XX	
XX	ABBS59631;
XX	
DT	26-MAR-2002 (first entry)
XX	
DE	Drosophila melanogaster polypeptide SEQ ID NO 5685.

XX Drosophila; developmental biology; cell signalling; insecticide;  
KM pharmaceutical.  
XX  
XX Drosophila melanogaster.  
XX  
XX WO200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
XX 23-MAR-2001; 2001WO-US09231.  
PF  
XX 23-MAR-2000; 2000US-191637P.  
PR 11-JUL-2000; 2000US-0614150.  
XX  
XX (PEKE ) PE CORP NY.  
XX  
XX Venter JC, Adams M, Li PWD, Myers EW;  
PI  
XX WPI; 2001-656860/75.  
DR  
XX N-PSDB; ABL03734.  
DR  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
PS  
XX Disclosure; SEQ ID NO 5685; 21bp + Sequence Listing; English.  
XX  
XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL01840-ABL16175) and the encoded proteins  
CC sequences (ABL01840-ABL16175) and the encoded proteins  
CC (ABB57737-ABB72072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 949 AA;  
SQ

Query Match 91.7%; Score 33; DB 22; Length 949;  
Best Local Similarity 85.7%; Pred. No. 2.2e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GQTRSPL 7  
ID ABB31116 standard; Peptide; 48 AA.  
DB 236 GETRSPL 242

RESULT 6  
ABB59856  
ID ABB59856 standard; Protein; 949 AA.  
XX  
XX ABB59856;  
AC  
XX  
XX 26-MAR-2002 (first entry)  
DT  
XX  
XX Drosophila melanogaster polypeptide SEQ ID NO 6360.  
DE  
XX Drosophila; developmental biology; cell signalling; insecticide;  
KM pharmaceutical.  
XX  
XX Drosophila melanogaster.  
OS  
XX WO200171042-A2.  
XX  
XX 27-SEP-2001.  
PD  
XX 23-MAR-2001; 2001WO-US09231.  
PF  
XX 23-MAR-2000; 2000US-191637P.  
PR  
XX 23-MAR-2000; 2000US-191637P.  
DR

PR 11-JUL-2000; 2000US-0614150.  
XX  
XX (PEKE ) PE CORP NY.  
XX  
XX Venter JC, Adams M, Li PWD, Myers EW;  
PI  
XX WPI; 2001-656860/75.  
DR  
XX N-PSDB; ABL03959.  
DR  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
PS  
XX Disclosure; SEQ ID NO 6360; 21bp + Sequence Listing; English.  
XX  
XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins  
CC (ABB57737-ABB72072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 949 AA;  
SQ

Query Match 91.7%; Score 33; DB 22; Length 949;  
Best Local Similarity 85.7%; Pred. No. 2.2e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GQTRSPL 7  
ID ABB31116 standard; Peptide; 48 AA.  
DB 236 GETRSPL 242

RESULT 7  
ABB31116  
ID ABB31116 standard; Peptide; 48 AA.  
XX  
XX ABB31116;  
AC  
XX  
XX 01-FEB-2002 (first entry)  
DT  
XX  
XX Peptide #3767 encoded by breast cell single exon nucleic acid probe.  
DE  
XX Human; microarray; single exon probe; gene expression; breast;  
KM disease; cancer.  
XX  
XX Homo sapiens.  
OS  
XX WO200157271-A2.  
XX  
XX 09-AUG-2001.  
PD  
XX  
XX 30-JAN-2001; 2001WO-US00662.  
PF  
XX  
XX 04-FEB-2000; 2000US-0180312.  
PR 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632365.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
PI  
XX WPI; 2001-496933/54.  
DR  
XX

PT New spatially-addressable set of single exon nucleic acid probes,  
 PT useful for measuring gene expression in sample derived from human  
 PT breast, comprises number of single exon nucleic acid probes -  
 XX  
 XX  
 PS Claim 27, SEQ ID NO 14084, 327pp + sequence listing; English.  
 CC The invention relates to a spatially-addressable set of single exon  
 CC nucleic acid probes for measuring gene expression in a sample derived  
 CC from human breast and BT 474 cells. The method involves contacting  
 CC the probes with a collection of detectably labelled nucleic acids  
 CC derived from mRNA of human breast, and then measuring the label  
 CC bound to each probe of the microarray. The probes are useful for  
 CC verifying the expression of regions of genomic DNA predicted to  
 CC encode proteins. They are useful for gene discovery, and for  
 CC determining predisposition and/or prognosing breast disease. Gene  
 CC expression analysis is useful for assessing the toxicity of chemical  
 CC agents on cells. The microarray of this invention presents a far greater  
 CC diversity of probes for measuring gene expression, with far less bias  
 CC than expressed sequence tag microarrays. The method is suitable for  
 CC rapid production of functional information from genomic sequence. The  
 CC present sequence is a peptide encoded by a single exon nucleic acid  
 CC probe of the invention.  
 CC Note: The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 XX  
 SO Sequence 48 AA;  
 Query Match 88.9%; Score 32; DB 22; Length 48;  
 Best Local Similarity 100.0%; Pred. No. 14;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 GQTRSP 6  
 Db 10 GQTRSP 15  
 DB 10 GQTRSP 15  
 RESULT 8  
 ABB36319  
 ID ABB36319 standard; Peptide; 48 AA.  
 AC ABB36319;  
 XX  
 XX  
 DT 04-FEB-2002 (first entry)  
 XX  
 DE Peptide #3825 encoded by human foetal liver single exon probe.  
 XX  
 KM Human, foetal liver; gene expression; single exon nucleic acid probe.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200157277-A2.  
 PD 09-AUG-2001.  
 XX  
 PF 30-JAN-2001; 2001MO-US006659.  
 XX  
 XX 04-FEB-2000; 2000US-0180312.  
 PR 26-MAY-2000; 2000US-0207456.  
 PR 30-JUN-2000; 2000US-0608408.  
 PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-0234687.  
 PR 27-SEP-2000; 2000US-0236359.  
 PR 04-OCT-2000; 2000GB-0024263.  
 XX  
 PA (MOLB-) MOLECULAR DYNAMICS INC.  
 XX  
 PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX  
 DR WPI; 2001-483447/52.  
 XX  
 PT Human genome-derived single exon nucleic acid probes useful for  
 PT analyzing gene expression in human fetal liver -

XX  
 PS Claim 27, SEQ ID NO 28954; 639pp + sequence listing; English.  
 XX  
 XX  
 CC The invention relates to a single exon nucleic acid probe for  
 CC measuring human gene expression in a sample derived from human foetal  
 CC liver. The single exon nucleic acid probes may be used for predicting,  
 CC measuring and displaying gene expression in samples derived from human  
 CC foetal liver. The present sequence is a peptide encoded by a single exon  
 CC nucleic acid probe of the invention.  
 CC Note: The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 XX  
 SO Sequence 48 AA;  
 Query Match 88.9%; Score 32; DB 22; Length 48;  
 Best Local Similarity 100.0%; Pred. No. 14;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 GQTRSP 6  
 Db 10 GQTRSP 15  
 DB 10 GQTRSP 15  
 RESULT 9  
 ABB21674  
 ID ABB21674 standard; Protein; 48 AA.  
 AC ABB21674;  
 XX  
 XX  
 DT 23-JAN-2002 (first entry)  
 XX  
 DE Protein #3673 encoded by probe for measuring heart cell gene expression.  
 XX  
 KM Human, gene expression; heart; microarray; vascular system;  
 KM cardiovascular disease; hypertension; cardiac arrhythmia;  
 KM congenital heart disease.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200157274-A2.  
 PD 09-AUG-2001.  
 XX  
 PF 30-JAN-2001; 2001MO-US00666.  
 XX  
 XX 04-FEB-2000; 2000US-0180312.  
 PR 26-MAY-2000; 2000US-0207456.  
 PR 30-JUN-2000; 2000US-0608408.  
 PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-0234687.  
 PR 27-SEP-2000; 2000US-0236359.  
 PR 04-OCT-2000; 2000GB-0024263.  
 XX  
 PA (MOLB-) MOLECULAR DYNAMICS INC.  
 XX  
 PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX  
 DR WPI; 2001-488899/53.  
 XX  
 PT Single exon nucleic acid probes for analyzing gene expression in human  
 PT hearts -  
 XX  
 PS Claim 15, SEQ ID NO 23444; 530pp; English.  
 XX  
 XX The present invention relates to single exon nucleic acid probes for  
 CC measuring human gene expression in a sample derived from human heart (see  
 CC ABA21535-ABA41305). The present sequence is a protein encoded by one such  
 CC probe. The probes may be used for predicting, measuring and displaying  
 CC gene expression in samples derived from the human heart via microarrays.  
 CC By measuring gene expression, the probes are useful for predicting,  
 CC diagnosing, grading, staging, monitoring and prognosing diseases of the  
 CC human heart and vascular system e.g. cardiovascular disease.

CC hypertension, cardiac arrhythmias and congenital heart disease.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 48 AA;

Query Match 88.9%; Score 32; DB 22; Length 48;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GQTRSP 6  
|||||  
Db 10 GQTRSP 15

RESULT 10  
AAM69471  
ID AAM69471 standard; Protein; 48 AA.

XX AAM69471;

DT 06-NOV-2001 (first entry)

DE Human bone marrow expressed probe encoded protein SEQ ID NO: 29777.

XX Human; bone marrow expressed exon; gene expression analysis; probe;  
KM microarray; cancer; leukemia; lymphoma; myeloma.

OS Homo sapiens.

XX WO200157276-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00668.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488900/53.

XX Human genome-derived single exon nucleic acid probes useful for  
PT analyzing gene expression in human bone marrow -

XX Example 4; SEQ ID NO: 29777; 658bp + Sequence Listing; English.

XX The present invention provides a number of single exon nucleic acid  
CC probes which are derived from genomic sequences expressed in the human  
CC bone marrow. They can be used to measure gene expression in bone marrow  
CC samples, which may enable the improved diagnosis and treatment of cancers  
CC such as lymphoma, leukemia and myeloma. The present sequence is a  
CC protein encoded by one of the probes of the invention.

XX Sequence 48 AA;

Query Match 88.9%; Score 32; DB 22; Length 48;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GQTRSP 6  
|||||  
Db 10 GQTRSP 15

RESULT 11  
AAM17314  
ID AAM17314 standard; Protein; 48 AA.

XX AAM17314;

DT 12-OCT-2001 (first entry)

DE Peptide #3748 encoded by probe for measuring cervical gene expression.

XX Probe; human; microarray; gene expression; cervical epithelial cell;  
KM cervical cancer.

OS Homo sapiens.

XX WO200157278-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00670.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488901/53.

XX Human genome-derived single exon nucleic acid probes useful for  
PT analyzing gene expression in human cervical epithelial cells -

XX Claim 27; SEQ ID NO 22140; 487bp; English.

XX The present invention relates to human single exon nucleic acid probes  
CC (SENPs; see AAI10068-AA128459). The present sequence is a peptide encoded  
CC by one such probe. The SENPs are derived from human HeLa cells. The SENPs  
CC can be used to produce a single exon microarray, which can be used for  
CC measuring human gene expression in a sample derived from human cervical  
CC epithelial cells. By measuring gene expression, the probes are therefore  
CC useful in grading and/or staging of diseases of the cervix, notably  
CC cervical cancer.

CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 48 AA;

Query Match 88.9%; Score 32; DB 22; Length 48;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GQTRSP 6  
|||||  
Db 10 GQTRSP 15

RESULT 12

AAM29812  
ID AAM29812 standard; Protein; 48 AA.

XX AAM29812;

XX 17-OCT-2001 (first entry)

XX Peptide #3849 encoded by probe for measuring placental gene expression.

```

XX KW Probe; microarray; human; placenta; antenatal diagnosis;
XX KW genetic disorder.
XX OS Homo sapiens.
XX PN WO200157272-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US00663.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236559.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX DR WPI; 2001-488897/53.
XX PT Human genome-derived single exon nucleic acid probes useful for
XX PT analyzing gene expression in human placenta -
XX PS Claim 27; SEQ ID No 30081; 654pp; English.
XX CC The present invention relates to single exon nucleic acid probes (SENP:
XX CC see A113135-A1157546). The present sequence is a microarray encoded by one
XX CC such probe. The probes are useful for producing a microarray for
XX CC predicting, measuring and displaying gene expression in samples derived
XX CC from human placenta. The probes are useful for antenatal diagnosis of
XX CC human genetic disorders.
XX SQ Sequence 48 AA;

Query Match      88.9%; Score 32; DB 22; Length 48;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GQTRSP 6
   |||||
Db 10 GQTRSP 15

RESULT 13
AAM04993
ID AAM04993 standard; Protein; 48 AA.
XX AC AAM04993;
XX DT 09-OCT-2001 (first entry)
XX DE Peptide #3675 encoded by probe for measuring breast gene expression.
XX KW Probe; human; breast disease; breast cancer; development disorder;
XX KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX OS Homo sapiens.
XX PN WO200157270-A2.
XX PD 09-AUG-2001.
XX PF 29-JAN-2001; 2001WO-US00661.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.

```

```

PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX DR WPI; 2001-476286/51.
XX PT Novel single exon nucleic acid probe used to measuring gene expression
XX PT in a human breast -
XX PS Claim 27; SEQ ID No 13733; 322pp; English.
XX CC The present invention relates to novel single exon nucleic acid probes
XX CC (see A1100010-A110067). The present sequence is a peptide encoded by one
XX CC such probe. The probes are useful for measuring human gene expression in
XX CC a human breast sample, where the probe hybridises at high stringency to a
XX CC nucleic acid expressed in the human breast. The probes are useful for
XX CC predicting, diagnosing, grading, staging, monitoring and prognosing
XX CC diseases of the human breast, particularly those diseases with polygenic
XX CC aetiology. The diseases include: breast cancer, disorders of development,
XX CC inflammatory diseases of the breast, fibrocystic changes, proliferative
XX CC breast disease and non-carcinoma tumours.
XX CC Note: The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 48 AA;

Query Match      88.9%; Score 32; DB 22; Length 48;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GQTRSP 6
   |||||
Db 10 GQTRSP 15

RESULT 14
ABG39102
ID ABG39102 standard; Peptide; 48 AA.
XX AC ABG39102;
XX DT 19-AUG-2002 (first entry)
XX DE Human peptide encoded by genome-derived single exon probe SEQ ID 28767.
XX KW Human; single exon probe; asthma; lung cancer; COPD; IHD;
XX KW chronic obstructive pulmonary disease; interstitial lung disease;
XX KW familial idiopathic pulmonary fibrosis; neurofibromatosis;
XX KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
XX KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosterosis;
XX KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karsenger syndrome;
XX KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
XX KW primary ciliary dyskinesia; pulmonary hypertension;
XX KW hyaline membrane disease.
XX OS Homo sapiens.
XX PN WO200186003-A2.
XX PD 15-NOV-2001.
XX PF 30-JAN-2001; 2001WO-US00665.
XX PR 04-FEB-2000; 2000US-180312P.
XX PR 26-MAY-2000; 2000US-207456P.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.

```





GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using bw model

Run on: January 29, 2003, 13:57:01 ; Search time 9.87179 Seconds  
(without alignments)  
68.168 Million cell updates/sec

Title: US-09-807-949a-109

Perfect score: 36

Sequence: 1 GQTRSPDL 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : PIR\_73:\*

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	32	88.9	144	2	EB4219
2	31	86.1	155	2	T15507
3	31	86.1	289	2	H70912
4	31	86.1	635	2	B64112
5	31	86.1	637	2	B83052
6	30	83.3	106	2	S51046
7	30	83.3	246	2	G72257
8	30	83.3	253	2	C84177
9	30	83.3	270	2	B83927
10	30	83.3	300	2	AB3188
11	30	83.3	302	1	RGE0MB
12	30	83.3	302	2	D91266
13	30	83.3	302	2	A86107
14	30	83.3	310	2	AT1022
15	30	83.3	405	2	B82643
16	30	83.3	518	2	G88961
17	30	83.3	899	2	S76449
18	30	83.3	1114	2	T49517
19	30	83.3	1806	2	AF1717
20	30	83.3	2139	2	A35672
21	29	80.6	125	2	F53086
22	29	80.6	135	2	S71852
23	29	80.6	213	1	I40569
24	29	80.6	213	1	S01734
25	29	80.6	213	1	S48126
26	29	80.6	336	2	B81210
27	29	80.6	341	2	C75436
28	29	80.6	408	2	B87436
29	29	80.6	416	2	G64542

30	29	80.6	422	2	C98233	glucose dehydrogen
31	29	80.6	437	2	JQ2015	nucleoprotein - ra
32	29	80.6	437	2	JQ2016	nucleoprotein - ra
33	29	80.6	450	1	VHYNRV	nucleoprotein - ra
34	29	80.6	450	1	VHYNRV	nucleoprotein - ra
35	29	80.6	450	1	VHYNRV	nucleoprotein - ra
36	29	80.6	450	2	A46104	nucleoprotein N -
37	29	80.6	580	2	G86169	hypothetical prote
38	29	80.6	635	2	B82273	dnak protein VC085
39	29	80.6	635	2	AT0057	chaperone protein
40	29	80.6	637	2	UC5608	dnak-type molecule
41	29	80.6	638	1	IQCCK	dnak-type molecule
42	29	80.6	638	2	P90630	heat shock protein
43	29	80.6	638	2	F85481	dnak-type molecule
44	29	80.6	638	2	AB0503	dnak protein (heat
45	29	80.6	656	1	S59631	endo-1,4-beta-xyla

#### ALIGNMENTS

RESULT 1  
E84219  
hypothetical protein Vng0613h [imported] - Halobacterium sp. NRC-1  
C/Species: Halobacterium sp. NRC-1  
C/Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #ext\_change 02-Feb-2001  
C/Accession: E84219  
R/NG, W.V.; Kennedy, S.P.; Mahatras, G.G.; Bergquist, B.; Pan, M.; Shukla, H.D.; Laaky, S.; Leitbauer, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jablor Jung, K.H.; Alam, M.; Freilich, T.  
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000  
A/Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebdardt, H.; Lowe, T.M.; Lié  
A/Title: Genome sequence of Halobacterium species NRC-1.  
A/Reference number: A84160; M01D:20504483; PMID:11016950  
A/Accession: E84219  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-144 <STO>  
A/Cross-references: GB:AE004437; NID:g10580207; PIDN:AA619121.1; GSPDB:GN00138  
C/Genetics:  
A/Gene: VNG0613H

Query Match 88.9%; Score 32; DB 2; Length 144;  
Best Local Similarity 85.7%; Pred. No. 9.5;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GQTRSPDL 7  
Db 134 GQTRDPL 140

RESULT 2  
T15507  
hypothetical protein C15C7.4 - Caenorhabditis elegans  
C/Species: Caenorhabditis elegans  
C/Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #ext\_change 04-Mar-2000  
C/Accession: T15507  
R/Leinbach, D.  
submitted to the EMBL Data Library, November 1995  
A/Description: The sequence of C. elegans cosmid C15C7.  
A/Reference number: Z18363  
A/Accession: T15507  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-155 <DEI>  
A/Cross-references: EMBL:U41528; NID:g1109795; PID:AA63152.1; CESP:C15C7  
C/Genetics:  
A/Gene: CESP:C15C7.4  
A/Intons: 16/2  
C/Superfamily: Caenorhabditis elegans hypothetical protein C15C7.4

Query Match 86.1%; Score 31; DB 2; Length 155;  
Best Local Similarity 85.7%; Pred. No. 17;

Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 GQTRSP1 7  
 |||||  
 Db 59 GQTRCPL 65

## RESULT 3

H70912  
 hypothetical protein Rv0048c - Mycobacterium tuberculosis (strain H37Rv)

C/Species: Mycobacterium tuberculosis  
 C/Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000  
 C/Accession: H70912  
 R/Cole, S.T.; Broese, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltham, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
 Nature 393, 537-544, 1998  
 A/Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrett, B.G.  
 A/Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
 A/Reference number: A70500; MUID:98295987; PMID:9634230  
 A/Accession: H70912  
 A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A/Molecule type: DNA  
 A/Residues: 1-289 <COL>  
 A/Cross-references: GB:280775; GB:AL123456; NID:93250715; PIDN:CAM02527.1; PID:91568587  
 A/Experimental source: strain H37Rv  
 C/Genetics:  
 A/Gene: Rv0048c  
 C/Superfamily: Mycobacterium tuberculosis hypothetical protein Rv0048c

Query Match  
 Best Local Similarity 86.1%; Score 31; DB 2; Length 289;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GQTRSP1 7  
 |||||  
 Db 112 GQTRSP1 118

## RESULT 4

B64112  
 dnak-type molecular chaperone - Haemophilus influenzae (strain Rd KW20)

N/Alternate names: heat shock protein 70  
 C/Species: Haemophilus influenzae  
 C/Date: 18-Aug-1995 #sequence\_revision 18-Aug-1995 #text\_change 20-Aug-1999  
 C/Accession: B64112  
 R/Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kervatage, P.; Godoyne, J.D.; Scott, J.; Shilley, R.; Liu, L.I.; Glodex, A.; Kelley, J.M.; Weidman, J.; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhmann, J.L.; Geoghegan, N.S.M.  
 Science 269, 496-512, 1995  
 A/Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, A/Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.  
 A/Reference number: A64000; MUID:95350630; PMID:7542800  
 A/Accession: B64112  
 A/Status: nucleic acid sequence not shown; translation not shown  
 A/Molecule type: DNA  
 A/Residues: 1-635 <TIGR>  
 A/Cross-references: GB:U32803; GB:U42023; NID:91574162; PIDN:AAC22889.1; PID:91574167; PID:91574167

C/Function:  
 A/Description: involved in protein folding and assembling/disassembling of protein comp  
 C/Superfamily: heat shock protein 70  
 C/Keywords: ATP; calcium; molecular chaperone; phosphoprotein

Query Match  
 Best Local Similarity 86.1%; Score 31; DB 2; Length 635;

Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GQTRSP1 7  
 |||||  
 Db 341 GQTRMPL 347

## RESULT 5

B83052  
 Dnak protein PA4761 (imported) - Pseudomonas aeruginosa (strain PA01)

C/Species: Pseudomonas aeruginosa  
 C/Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
 C/Accession: B83052  
 R/Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Bro  
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,  
 .; Lory, S.; Olson, M.V.  
 Nature 406, 959-964, 2000  
 A/Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho  
 A/Reference number: A82950; MUID:20437337; PMID:10984043  
 A/Accession: B83052  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-637 <STO>  
 A/Cross-references: GB:AE004889; GB:AE004091; NID:99951014; PIDN:AMG08147.1; GSPDB:GN001  
 A/Experimental source: strain PA01  
 C/Genetics:  
 A/Gene: dnak; PA4761  
 C/Superfamily: heat shock protein 70

Query Match  
 Best Local Similarity 86.1%; Score 31; DB 2; Length 637;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GQTRSP1 7  
 |||||  
 Db 342 GQTRMPL 348

## RESULT 6

S51046  
 hypothetical protein 1 - Paracoccus denitrificans

C/Species: Paracoccus denitrificans  
 C/Date: 01-Aug-1995 #sequence\_revision 01-Sep-1995 #text\_change 08-Oct-1999  
 C/Accession: S51046  
 R/van Spanning, R.J.M.; van der Palen, C.J.N.M.; Slotboom, D.J.; Reijnders, W.N.M.; Stout  
 Eur. J. Biochem. 226, 201-210, 1994  
 A/Title: Expression of the mau genes involved in methylamine metabolism in Paracoccus der  
 A/Reference number: S51046; MUID:95045590; PMID:7957249  
 A/Accession: S51046  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-106 <VAN>  
 A/Cross-references: EMBL:U12464; NID:9558801; PIDN:AAA56721.1; PID:9558802

Query Match  
 Best Local Similarity 83.3%; Score 30; DB 2; Length 106;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 GQTRSP1 7  
 |||||  
 Db 42 GQTRAPV 48

## RESULT 7

G72257  
 hypothetical protein - Thermococcus maritima (strain MSB8)

C/Species: Thermococcus maritima  
 C/Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 28-Jul-2000  
 C/Accession: G72257  
 R/Nelson, K.E.; Clayton, R.A.; Gail, S.R.; Gwin, M.L.; Dodson, R.J.; Haft, D.H.; Hickey,  
 Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.D.;  
 C.M.

Nature 399, 323-329, 1999  
 A/Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq  
 A/Reference number: A72200; MUID:99287316; PMID:10360571

A/Accession: G72257  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-246 <ARN>  
 A/Cross-references: GB:AE001793; GB:AE000512; NID:94981963; PIDN:AAD36477.1; PID:94981977  
 A/Experimental source: strain MSB8

C/Genetics:  
A/Gene: TM1406  
C/Superfamily: Thermotoga maritima hypothetical protein TM1406

Query Match 83.3%; Score 30; DB 2; Length 246;  
Best Local Similarity 100.0%; Pred. No. 46;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 QTRSP 7  
Db 202 QTRSP 207

RESULT 8  
C84177  
proteasome, subunit beta [imported] - Halobacterium sp. NRC-1

C/Species: Halobacterium sp. NRC-1  
C/Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 16-Feb-2001  
C/Accession: C84177  
R/NG: W.V.; Kennedy, S.P.; Mahairas, G.G.; Bergquist, B.; Pan, M.; Shukla, H.D.; Laeky, S.; Leitbauer, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jabcic, Jung, K.H.; Alam, M.; Freilias, T.  
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000

A/Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Li, A.; Title: Genome sequence of Halobacterium species NRC-1.  
A/Reference number: A84160; WUID:20504483; PMID:11016950

A/Accession: C84177  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-253 <STO>  
A/Cross-references: GB:AE004437; NID:G10579815; PIDN:AAQ18783.1; GSPDB:GN00138

C/Genetics:  
A/Gene: pamb  
C/Superfamily: multicatalytic endopeptidase complex chain C9

Query Match 83.3%; Score 30; DB 2; Length 253;  
Best Local Similarity 100.0%; Pred. No. 48;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 QTRSP 7  
Db 55 QTRSP 60

RESULT 9

B83927  
hypothetical protein B83927 [imported] - Bacillus halodurans (strain C-125)

C/Species: Bacillus halodurans  
C/Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 15-Jun-2001

A/Accession: B83927  
R/Nakani, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; Hira, Nucleic Acids Res. 28, 4317-4331, 2000

A/Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and A/Reference number: A83650; WUID:20512582; PMID:11058132

A/Accession: B83927  
A/Status: preliminary  
A/Molecule type: DNA

A/Residues: 1-270 <STO>  
A/Cross-references: GB:AP001514; GB:BA000004; NID:G10174613; PIDN:BA05937.1; GSPDB:GN004  
A/Experimental source: strain C-125  
C/Genetics:

A/Gene: B82218

Query Match 83.3%; Score 30; DB 2; Length 270;  
Best Local Similarity 100.0%; Pred. No. 51;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 QTRSP 7  
Db 120 QTRSP 125

RESULT 10

AB3188  
acetyltransferase [imported] - Agrobacterium tumefaciens (strain C58, Dupont) plasmid AT

C/Species: Agrobacterium tumefaciens  
C/Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 11-Jan-2002

A/Accession: AB3188  
R/Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.

erage, G.; Gillet, W.; Grant, C.; Genthner, D.; Kutyavin, T.; Levy, R.; Li, M.; McElle, Science 294, 2317-2323, 2001

A/Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, E ster, B.W.

A/Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
A/Reference number: AB2577; PMID:11743193

A/Accession: AB3188  
A/Status: preliminary  
A/Molecule type: DNA

A/Residues: 1-300 <KUR>  
A/Cross-references: GB:AE008687; PIDN:AA45920.1; PID:G17743667; GSPDB:GN00188

A/Experimental source: strain C58 (Dupont)  
C/Genetics:  
A/Gene: Acus231  
A/Genome: plasmid

Query Match 83.3%; Score 30; DB 2; Length 300;  
Best Local Similarity 71.4%; Pred. No. 57;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 QTRSP 7  
Db 284 QTRSP 290

RESULT 11

RGBCMB  
melibiose operon regulatory protein - Escherichia coli (strain K-12)

C/Species: Escherichia coli  
C/Date: 31-Dec-1988 #sequence\_revision 31-Dec-1988 #text\_change 01-Mar-2002

A/Accession: A29625; S56347; B65221  
R/Webster, C.; Kempell, K.; Booth, I.; Bueby, S.

Gene 59, 253-263, 1987  
A/Title: Organisation of the regulatory region of the Escherichia coli melibiose operon.  
A/Reference number: A29625; WUID:88337961; PMID:2830169

A/Accession: A29625  
A/Molecule type: DNA

A/Residues: 1-302 <WEB>  
R/Burand, V.; Plunkett III, G.; Sofia, H.J.; Daniels, D.L.; Blattner, F.R.

Nucleic Acids Res. 23, 2105-2119, 1995  
A/Title: Analysis of the Escherichia coli genome VI: DNA sequence of the region from 92.f  
A/Reference number: S56314; WUID:95334362; PMID:7610040

A/Accession: S56347  
A/Status: preliminary; nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA  
A/Residues: 1-302 <BUR>

A/Cross-references: EMBL:U14003; NID:G1263172; PIDN:AAA97018.1; PID:G536963  
A/Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1994

R/Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burand, V.; Riley, M.; Col A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997  
A/Title: The complete genome sequence of Escherichia coli K-12.  
A/Reference number: A64720; WUID:97426617; PMID:9278503

A/Accession: B65221  
A/Status: preliminary; nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA  
A/Residues: 1-302 <BLAT>

A/Cross-references: GB:AE000484; GB:U00096; NID:G2367352; PIDN:AACT7079.1; PID:G1790559;  
A/Experimental source: strain K-12, substrain MG1655  
C/Genetics:

A/Gene: melR  
A/Map position: 93 min

C/Function:  
A/Description: may be a melibiose-induced activator for the melibiose operon promoter or et define

C/Superfamily: arabinose operon regulatory protein

C;Keywords: DNA binding; transcription regulation

Query Match 83.3%; Score 30; DB 1; Length 302;  
Best Local Similarity 100.0%; Pred. No. 58;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 QTRSP 7  
Db 14 QTRSP 19

RESULT 12  
D91266

regulator of melibiose operon [imported] - *Escherichia coli* (strain O157:H7, substrain R  
C;Species: *Escherichia coli*

C;Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 24-Aug-2001

C;Accession: D91266

R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.

gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.

DNA Res. 8, 11-22, 2001

A;Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and gene

A;Reference number: A89629; MUID:21156231; PMID:11258796

A;Accession: D91266

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-302 <HAY>

A;Cross-references: GB:BA000007; PIDN:BA838523.1; PID:G13364577; GSPDB:GN00154

A;Experimental source: strain O157:H7, substrain R1MD 0509952

C;Genetics:

A;Gene: EC65100

C;Superfamily: arabinose operon regulatory protein

Query Match 83.3%; Score 30; DB 2; Length 302;  
Best Local Similarity 100.0%; Pred. No. 58;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 QTRSP 7  
Db 14 QTRSP 19

RESULT 13  
A86107

regulator of melibiose operon [imported] - *Escherichia coli* (strain O157:H7, substrain R  
C;Species: *Escherichia coli*

C;Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001

C;Accession: A86107

R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew

Miller, L.; Grobeck, E.J.; Davis, N.W.; Lam, A.; Dimalanta, E.; Potamousis, K.; Apodaca,

Nature 409, 529-533, 2001

A;Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.

A;Reference number: A85460; MUID:21074935; PMID:11206551

A;Accession: A86107

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-302 <STO>

A;Cross-references: GB:AA005174; NID:G12519091; PIDN:AA059317.1; GSPDB:GN00145; UMGF:257

A;Experimental source: strain O157:H7, substrain EDU933

C;Genetics:

A;Gene: melR

C;Superfamily: arabinose operon regulatory protein

Query Match 83.3%; Score 30; DB 2; Length 302;  
Best Local Similarity 100.0%; Pred. No. 58;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 QTRSP 7  
Db 14 QTRSP 19

RESULT 14  
A11022

melibiose operon regulatory protein melR [imported] - *Salmonella enterica* subsp. *enterica*  
C;Species: *Salmonella enterica* subsp. *enterica* serovar Typh

A;Note: this species has also been called *Salmonella typhi*

C;Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 27-Nov-2001

C;Accession: A11022

R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,

Ch, T.; Conerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,

S.; Mout, S.; O'Garra, P.

Nature 413, 848-852, 2001

A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skellon, J.; Stevens, K.

A;Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* serov

A;Reference number: AB0502; PMID:11677608

A;Accession: A11022

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-310 <PAR>

A;Cross-references: GB:AL513382; PIDN:CAD09282.1; PID:G16505286; GSPDB:GN00176

C;Genetics:

A;Gene: melR

C;Superfamily: arabinose operon regulatory protein

Query Match 83.3%; Score 30; DB 2; Length 310;  
Best Local Similarity 100.0%; Pred. No. 59;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 QTRSP 7  
Db 22 QTRSP 27

RESULT 15  
B82643  
glucose dehydrogenase B XP1740 [imported] - *Xylella fastidiosa* (strain 9asc)

C;Species: *Xylella fastidiosa*

C;Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Sep-2000

C;Accession: B82643

R;anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequenc

Nature 406, 151-157, 2000

A;Title: The genome sequence of the plant pathogen *Xylella fastidiosa*.

A;Reference number: A82515; MUID:20365717; PMID:10910347

A;Note: for a complete list of authors see reference number A59328 below

A;Accession: B82643

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-405 <STM>

A;Cross-references: GB:AE003997; GB:AE003849; NID:G9106805; PIDN:AA084549.1; GSPDB:GN001

A;Experimental source: strain 9asc

R;Simpson, A.J.G.; Reinach, P.C.; Artuda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A

Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Carraro, D.M.; Carter, H

as-Neto, E.; Docena, C.; El-Dorri, H.; Facincent, A.P.; Ferreira, A.J.S.

submitted to GenBank, June 2000

A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; France, S.C.; Franco, M.C.; Frohme

U.D.; Ungerleir, M.L.; Kemper, E.L.; Kitejima, J.P.; Krueger, J.B.; Kuramae, E.E.; Laigre

chado, M.A.; Madelira, A.M.B.N.; Madelira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E

A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;

F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A

Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Savasak

A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir

M.; Teshako, M.H.; Valleda, H.; Van Sluys, M.A.; Verjovskij-Almeida, S.; Vettore, A.L.; Z

A;Reference number: A59328

A;Contents: annotation

C;Genetics:

A;Gene: XP1740

C;Superfamily: hypothetical protein b0837

Query Match 83.3%; Score 30; DB 2; Length 405;  
Best Local Similarity 71.4%; Pred. No. 78;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QTRSP 7  
Db 49 QTRSP 55

Thu Jan 30 11:47:41 2003

us-09-807-949a-109.rpr

Page 5

Search completed: January 29, 2003, 14:03:25  
Job time : 10.8718 secs

---

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Comogen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2003, 13:51:40 ; Search time 4.84615 Seconds  
(without alignments)  
59.910 Million cell updates/sec

Title: US-09-807-949a-109  
Perfect score: 36  
Sequence: 1 GQTRSP 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues  
Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	31	86.1	289	Y048_MYCTU	P71705 mycobacteri
2	31	86.1	631	DNAX_PASHA	O52064 pasteurella
3	31	86.1	631	DNAX_THERO	P96133 thermomicro
4	31	86.1	632	DNAX_ACTAC	P71331 actinobacil
5	31	86.1	633	DNAX_HAEDU	P48209 haemophilus
6	31	86.1	634	DNAX_HAEDU	P43736 haemophilus
7	31	86.1	634	DNAX_PASMU	P57870 pasteurella
8	31	86.1	637	DNAX_PESAS	O9hw43 pseudomonas
9	31	86.1	638	DNAX_PESAS	O9hw43 pseudomonas
10	31	86.1	642	DNAX_PASMU	P48205 franciella
11	31	86.1	644	DNAX_PASMU	O32482 legionella
12	31	86.1	866	AREA_ASPOR	O13415 aspergillus
13	30	83.3	253	PSMA_HALNI	P57697 halobacteri
14	30	83.3	302	MEIR_ECOLI	P10411 escherichia
15	30	83.3	2139	CRB_DROME	P10040 drosophila
16	29	80.6	213	XYNA_BACCI	P09860 bacillus ci
17	29	80.6	213	XYNA_BACCI	P18423 bacillus su
18	29	80.6	416	GLYA_HAEPY	P56089 helicobacte
19	29	80.6	450	NCAP_RABVA	P15197 rabies viru
20	29	80.6	450	NCAP_RABVA	O08314 rabies viru
21	29	80.6	450	NCAP_RABVA	P06025 rabies viru
22	29	80.6	450	NCAP_RABVA	P16285 rabies viru
23	29	80.6	450	NCAP_RABVA	O09110 rabies viru
24	29	80.6	542	MODU_DROME	P13469 drosophila
25	29	80.6	635	DNAX_VIBCH	O34241 vibrio chol
26	29	80.6	636	DNAX_YERPE	O82107 yersinia pe
27	29	80.6	637	DNAX_BUCAI	O32464 buchiera ap
28	29	80.6	637	DNAX_ECOLI	P04475 escherichia
29	29	80.6	637	DNAX_SALTI	O82911 salmonella
30	29	80.6	637	DNAX_SALTY	O56073 salmonella
31	29	80.6	640	DNAX_VIBPR	O91721 vibrio proc
32	29	80.6	640	DNAX_VIBHA	O87384 vibrio harv
33	29	80.6	936	MSHA_HUMAN	O15457 homo sapien

34	28	77.8	133	1	RS8_SYNY3	P73307 synchocyst
35	28	77.8	134	1	YN82_YEAST	P53717 saccharomyc
36	28	77.8	177	1	PI11_HUMAN	O14683 homo sapien
37	28	77.8	198	1	SORC_CRITIO	P05044 cricetulus
38	28	77.8	198	1	SORC_HUMAN	P30826 homo sapien
39	28	77.8	208	1	XLR_MOUSE	P05531 mus musculu
40	28	77.8	356	1	CYS3_LYCERS	O40143 lycopersico
41	28	77.8	364	1	FLGI_ZYMOO	O92519 zymomonas m
42	28	77.8	384	1	FLGI_ZYMOO	P49936 bacillus su
43	28	77.8	393	1	CRFP_RHOCA	P17061 rhodobacter
44	28	77.8	417	1	YEIM_HAEDU	P44742 haemophilus
45	28	77.8	421	1	HMDH_AERPE	O9y484 aeropyrum p

## ALIGNMENTS

RESULT 1  
Y048\_MYCTU STANDARD; PRT; 289 AA.  
ID Y048\_MYCTU  
AC P71705;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 15-UTN-2002 (Rel. 41, Last annotation update)  
DE Hypothetical protein RV0048c precursor.  
GN RV0048C OR MT0054 OR MTCY21D4.11C.  
OS Mycobacterium tuberculosis  
OC Bacteria; Actinobacteria; Actinobacteridae (class); Actinobacteriales; Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_Taxid:1773;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=H37Rv;  
RX MEDLINE=98295987; PubMed=9634230;  
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D., Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F., Badcock K., Baeham D., Brown D., Chillingworth T., Connor R., Davies R., Devlin K., Feldwell T., Gentles S., Hamlin N., Holtroyd S., Hornsby T., Jorgels K., Krogh A., McLean J., Moule S., Murphy L., Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J., Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;  
RA "Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence."  
RL Nature 393:537-544(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CDC 1551 / Oshkosh;  
RA Fiechtmann R.D., Alland D., Eisen J.A., Carpenter L., White O., Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickery E., Kolonay J.F., Nelson W.C., Umayam L.A., Rasmussen M.D., Salzberg S.L., Delcher A.L., Utterback T., Weidman J., Khouri H., Gill J., Mikula A., Bishai W.;  
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and laboratory strains."  
RT Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.  
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
CC -----  
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: Z80775; CAB02527.1; -;  
DR EMBL: AE006918; AKK44276.1; -;  
DR TIGR: MT0054; -;  
DR Tuberculist: RV0048c; -;  
KW Hypothetical protein; Transmembrane; Signal; Complete proteome.  
FT SIGNAL 1 19  
FT CHAIN 20 289 HYPOTHETICAL PROTEIN RV0048C.

FT TRANSMEM 90 110 POTENTIAL.  
 FT TRANSMEM 257 277 POTENTIAL.  
 FT CONFLICT 248 248 E -> D (IN REF. 2).  
 FT CONFLICT 250 250 V -> A (IN REF. 2).  
 SQ SEQUENCE 289 AA; 30817 MW; C2FC916EDC7711B CRC64;

Query Match 86.1%; Score 31; DB 1; Length 289;  
 Best Local Similarity 85.7%; Pred. No. 12;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GQTRSP 7  
 DB 112 GQTRSP 118

## RESULT 2

DNK PASHA STANDARD; PRT; 631 AA.  
 ID DNK PASHA STANDARD; PRT; 631 AA.  
 AC 052064;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Chapterone protein dnak (Heat shock protein 70) (Heat shock 70 kDa protein) (HSP70).  
 GN DNK.  
 OS Pasteurella haemolytica.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
 OC Mannheimia.  
 OC NCBI\_TaxID=75985;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Serotype A1;  
 RA Al S. L., Lo R.Y. C.;  
 RL "The dnak and dnaJ chapterone genes of Pasteurella haemolytica A1," submitted (Aug-1997) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: ACTS AS A CHAPERONE (BY SIMILARITY).  
 CC -1- INDUCTION: BY STRESS CONDITIONS E.G. HEAT SHOCK (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AF017730; AAB94554.1; -;  
 DR HSSP; P04475; 1DG4.  
 DR InterPro: IPR001023; Hsp70.  
 DR Pfam: PF000012; HSP70; 1.  
 DR PRINTS; PR00301; HEATSHOCK70.  
 DR ProDom; PD000089; Hsp70; 1.  
 DR PROSITE; PS00297; HSP70\_1; 1.  
 DR PROSITE; PS00329; HSP70\_2; 1.  
 DR PROSITE; PS01036; HSP70\_3; 1.  
 KW Chapterone; ATP-binding; Heat shock; Phosphorylation.  
 FT MOD\_RSS 198 198 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 SQ SEQUENCE 631 AA; 67979 MW; 4DE9FEC3A36B4B6F CRC64;

Query Match 86.1%; Score 31; DB 1; Length 631;  
 Best Local Similarity 85.7%; Pred. No. 28;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GQTRSP 7  
 DB 340 GQTRSP 346

RESULT 3  
 DNK THERO STANDARD; PRT; 631 AA.  
 ID DNK THERO STANDARD; PRT; 631 AA.  
 AC P96133; 1

DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Chapterone protein dnak (Heat shock protein 70) (Heat shock 70 kDa protein) (HSP70).  
 GN DNK.  
 OS Thermomicrobium roseum.  
 OC Bacteria; Thermomicrobia; Thermomicrobia (class); Thermomicrobiales;  
 OC Thermomicrobiaceae; Thermomicrobium.  
 OC NCBI\_TaxID=500;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 27502;  
 RX MEDLINE=97144518; PubMed=8990285;  
 RA Gupta R.S., Bustard K., Falah M., Singh D.;  
 RT "Sequencing of heat shock protein 70 (dnak) homologs from Deinococcus proteolyticus and Thermomicrobium roseum and their integration in a protein-based phylogeny of prokaryotes.";  
 RL J. Bacteriol. 179:345-357(1997).  
 CC -1- FUNCTION: ACTS AS A CHAPERONE (BY SIMILARITY).  
 CC -1- INDUCTION: BY STRESS CONDITIONS E.G. HEAT SHOCK (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; U80216; AAB41740.1; -;  
 DR HSSP; P04475; 1DG4.  
 DR InterPro: IPR001023; Hsp70.  
 DR Pfam; PF00012; HSP70; 1.  
 DR PRINTS; PR00301; HEATSHOCK70.  
 DR ProDom; PD000089; Hsp70; 1.  
 DR PROSITE; PS00297; HSP70\_1; 1.  
 DR PROSITE; PS00329; HSP70\_2; 1.  
 DR PROSITE; PS01036; HSP70\_3; 1.  
 KW Chapterone; ATP-binding; Heat shock; Phosphorylation.  
 FT MOD\_RSS 197 197 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 SQ SEQUENCE 631 AA; 68922 MW; 1BA4056B2EA13DBE CRC64;

Query Match 86.1%; Score 31; DB 1; Length 631;  
 Best Local Similarity 85.7%; Pred. No. 28;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GQTRSP 7  
 DB 336 GQTRSP 342

RESULT 4  
 DNK ACTAC STANDARD; PRT; 632 AA.  
 ID DNK ACTAC STANDARD; PRT; 632 AA.  
 AC P11331;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Chapterone protein dnak (Heat shock protein 70) (Heat shock 70 kDa protein) (HSP70).  
 GN DNK.  
 OS Actinobacillus actinomycetemcomitans (Haemophilus actinomycetemcomitans).  
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
 OC Actinobacillus.  
 OC NCBI\_TaxID=714;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Y4;  
 RA Yoshida A., Nakano Y., Yamashita Y., Yu H., Ohishi M., Koga T.;  
 RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.



CC -1- FUNCTION: ACTS AS A CHAPERONE (BY SIMILARITY).  
 CC -1- INDUCTION: BY STRESS CONDITIONS E.G. HEAT SHOCK (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL; D87753; BAA13454.1; -  
 DR HSSP; P04475; 1DG4.  
 DR InterPro: IPR001023; Hsp70.  
 DR Pfam: PF00012; HSP70\_1.  
 DR PRINTS; PR00301; HEATSHOCK70.  
 DR ProDom; PD000089; HSP70\_2.  
 DR PROSITE; PS00297; HSP70\_1; 1.  
 DR PROSITE; PS00329; HSP70\_2; 1.  
 DR PROSITE; PS01036; HSP70\_3; 1.  
 KM Chaperone; ATP-binding; Heat shock; Phosphorylation.  
 FT INIT MET 0  
 FT MOD RES 198 198 BY SIMILARITY.  
 FT MOD RES 198 198 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 SQ SEQUENCE 632 AA; 68263 MW; B2CC46687B8235A6 CRC64;  
 Query Match  
 Best Local Similarity 86.1%; Score 31; DB 1; Length 632;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 GQTRSP 7  
 DB 341 GQTRMPL 347  
 RESULT 5  
 DNAX HAEDU STANDARD; PRT; 633 AA.  
 AC P48209;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Chaperone protein dnax (Heat shock protein 70) (Heat shock 70 kDa  
 DE protein) (HSP70).  
 GN DNAX.  
 OS Haemophilus ducreyi.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
 OC Haemophilus.  
 OX NCBI\_TaxId=730;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=35000;  
 RL Parsons L.M.;  
 RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: ACTS AS A CHAPERONE (BY SIMILARITY).  
 CC -1- INDUCTION: BY STRESS CONDITIONS E.G. HEAT SHOCK (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL; U2596; AAA67298.1; -  
 DR HSSP; P04475; 1DG4.  
 DR InterPro: IPR001023; Hsp70.  
 DR Pfam; PF00012; HSP70\_1.  
 DR PRINTS; PR00301; HEATSHOCK70.  
 DR ProDom; PD000089; Hsp70\_1.  
 DR PROSITE; PS00297; HSP70\_1; 1.  
 DR PROSITE; PS01036; HSP70\_3; 1.  
 KM Chaperone; ATP-binding; Heat shock; Phosphorylation;  
 FT INIT MET 0  
 FT MOD RES 197 197 BY SIMILARITY.  
 FT MOD RES 197 197 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 SQ SEQUENCE 634 AA; 68149 MW; 4892B31B2EC3CC9 CRC64;

DR PROSITE; PS00329; HSP70\_2; 1.  
 DR PROSITE; PS01036; HSP70\_3; 1.  
 KM Chaperone; ATP-binding; Heat shock; Phosphorylation.  
 FT INIT MET 0  
 FT MOD RES 197 197 BY SIMILARITY.  
 FT MOD RES 197 197 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 SQ SEQUENCE 633 AA; 68418 MW; 87D232B060600EBC CRC64;  
 Query Match  
 Best Local Similarity 86.1%; Score 31; DB 1; Length 633;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 GQTRSP 7  
 DB 340 GQTRMPL 346  
 RESULT 6  
 DNAX HAEDU STANDARD; PRT; 634 AA.  
 AC P43736;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Chaperone protein dnax (Heat shock protein 70) (Heat shock 70 kDa  
 DE protein) (HSP70).  
 GN DNAX OR H1237.  
 OS Haemophilus influenzae.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
 OC Haemophilus.  
 OX NCBI\_TaxId=727;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Rd / KW20 / ATCC 51907;  
 RX MEDLINE=95350630; PubMed=7542800;  
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
 RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,  
 RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,  
 RA Scott J.D., Shiley R., Liu L.-I., Glodek A., Kelley J.M.,  
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,  
 RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudak D.M., Brandon R.C.,  
 RA Fine L.D., Fritchman J.L., Fritchman J.L., Geoghegan N.S.M.,  
 RA Ghem C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  
 RA Venter J.C.;  
 RT "Whole-genome random sequencing and assembly of Haemophilus influenzae  
 RT Rd".  
 RL Science 269:496-512(1995).  
 CC -1- FUNCTION: ACTS AS A CHAPERONE (BY SIMILARITY).  
 CC -1- INDUCTION: BY STRESS CONDITIONS E.G. HEAT SHOCK (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL; U32803; AAC22889.1; -  
 DR HSSP; P04475; 1DG4.  
 DR TIGR; H1237; -  
 DR InterPro: IPR001023; Hsp70.  
 DR Pfam; PF00012; HSP70\_1.  
 DR PRINTS; PR00301; HEATSHOCK70.  
 DR ProDom; PD000089; Hsp70\_1.  
 DR PROSITE; PS00297; HSP70\_1; 1.  
 DR PROSITE; PS00329; HSP70\_2; 1.  
 DR PROSITE; PS01036; HSP70\_3; 1.  
 KM Chaperone; ATP-binding; Heat shock; Phosphorylation;  
 FT INIT MET 0  
 FT MOD RES 197 197 BY SIMILARITY.  
 FT MOD RES 197 197 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 SQ SEQUENCE 634 AA; 68149 MW; 4892B31B2EC3CC9 CRC64;

Query Match 86.1%; Score 31; DB 1; Length 634;  
 Best Local Similarity 85.7%; Pred. No. 28;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GQTRSP1 7  
 DB 340 GQTRMPL 346

## RESULT 7

DNAX\_PASNU STANDARD; PRT; 634 AA.  
 ID DNAX\_PASNU STANDARD; PRT; 634 AA.  
 AC P57870;  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Chaperrone protein dnax (Heat shock protein 70) (Heat shock 70 kDa protein) (HSP70).  
 GN DNAX OR PM0736.  
 OS Pasteurella multocida.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
 CC Pasteurella.  
 CC NCBI\_TaxID=747;  
 RX MEDLINE=2114586; PubMed=11248100;  
 RA May B.J., Zhang Q., Li L., Paustian M.L., Whitam T.S., Kapur V.,  
 RT "Complete genomic sequence of Pasteurella multocida pm70.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).  
 CC -1- FUNCTION: ACTS AS A CHAPERONE (BY SIMILARITY).  
 CC -1- INDUCTION: BY STRESS CONDITIONS E.G. HEAT SHOCK (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AEO06111; AKO2820.1; -;  
 DR HSSP; P04475; IDG4.  
 DR InterPro; IPR001023; Hsp70.  
 DR Pfam; PF00012; HSP70; 1.  
 DR PRINTS; PR00301; HEATSHOCK70.  
 DR ProDom; PD000089; Hsp70; 1.  
 DR PROSITE; PS00297; HSP70\_1; 1.  
 DR PROSITE; PS00329; HSP70\_2; 1.  
 DR PROSITE; PS01036; HSP70\_3; 1.  
 DR PROSITE; PS01036; HSP70\_3; 1.  
 KW Chaperrone; ATP-binding; Heat shock; Phosphorylation;  
 KM Complete proteome.  
 FT MOD RES 199 199 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 SQ SEQUENCE 634 AA; 68431 MM; 060F34200E5ACB20 CRC64;

Query Match 86.1%; Score 31; DB 1; Length 634;  
 Best Local Similarity 85.7%; Pred. No. 28;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GQTRSP1 7  
 DB 342 GQTRMPL 348

## RESULT 8

DNAX\_PSEAB STANDARD; PRT; 637 AA.  
 ID DNAX\_PSEAB STANDARD; PRT; 637 AA.  
 AC Q9HVA3;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Chaperrone protein dnax (Heat shock protein 70) (Heat shock 70 kDa protein) (HSP70).  
 GN DNAX OR PA4761.  
 OS Pseudomonas aeruginosa.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
 CC Pseudomonas.  
 CC NCBI\_TaxID=287;  
 RX MEDLINE=20437337; PubMed=10984043;  
 RA Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warren P.,  
 RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,  
 RA Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,  
 RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,  
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,  
 RA Reizer U., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;  
 RT "Complete genome sequence of Pseudomonas aeruginosa PAO1, an  
 RT opportunistic pathogen.";  
 RL Nature 406:959-964(2000).

CC -1- INDUCTION: ACTS AS A CHAPERONE (BY SIMILARITY).  
 CC -1- INDUCTION: BY STRESS CONDITIONS E.G. HEAT SHOCK (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AEO04889; AAG08147.1; -;  
 DR HSSP; P04475; IDG4.  
 DR InterPro; IPR001023; Hsp70.  
 DR Pfam; PF00012; HSP70; 1.  
 DR PRINTS; PR00301; HEATSHOCK70.  
 DR ProDom; PD000089; Hsp70; 1.  
 DR PROSITE; PS00297; HSP70\_1; 1.  
 DR PROSITE; PS00329; HSP70\_2; 1.  
 DR PROSITE; PS01036; HSP70\_3; 1.  
 DR PROSITE; PS01036; HSP70\_3; 1.  
 KW Chaperrone; ATP-binding; Heat shock; Phosphorylation;  
 KM Complete proteome.  
 FT MOD RES 199 199 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 SQ SEQUENCE 637 AA; 66403 MM; 06EA32E906486A9 CRC64;

Query Match 86.1%; Score 31; DB 1; Length 637;  
 Best Local Similarity 85.7%; Pred. No. 28;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GQTRSP1 7  
 DB 342 GQTRMPL 348

## RESULT 9

DNAX\_PSESG STANDARD; PRT; 638 AA.  
 ID DNAX\_PSESG STANDARD; PRT; 638 AA.  
 AC Q9WVG9;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Chaperrone protein dnax (Heat shock protein 70) (Heat shock 70 kDa protein) (HSP70).  
 GN DNAX.  
 OS Pseudomonas syringae (pv. glyciniae).  
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
 CC Pseudomonas.  
 CC NCBI\_TaxID=313;  
 RX MEDLINE=99407915; PubMed=10478477;

```

RA Keith L.M.W., Partridge J.E., Bender C.L.;
RT "dnak and the heat stress response of Pseudomonas syringae pv.
RT glycinea.";
RL Mol. Plant Microbe Interact. 12:563-574(1999).
CC -1- FUNCTION: ACTS AS A CHAPERONE (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF135163; AAD31868.1; -.
CC HSSP: P04475; 1DKY.
CC InterPro: IPR001023; Hsp70.
CC Pfam: PF00012; HSP70; 1.
CC PRINTS: PR00301; HEATSHOCK70.
CC ProDom: PD000089; HSP70; 1.
CC PROSITE: PS00297; HSP70_1; 1.
CC PROSITE: PS00329; HSP70_2; 1.
CC PROSITE: PS01036; HSP70_3; 1.
CC Chapterone; ATP-binding; Heat shock; Phosphorylation.
CC MOD RES 199 199 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
CC SEQUENCE 638 AA; 68897 MW; DE2C34D28BFRC1B CRC64;
CC -----
Query Match 86.1%; Score 31; DB 1; Length 638;
Best Local Similarity 85.7%; Pred. No. 28;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 GQTRSP 7
DB 342 GQTRMPL 348
-----
RESULT 10
DNAX FRATU STANDARD; PRT; 642 AA.
AC P48205;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Chapterone protein dnak (Heat shock protein 70) (Heat shock 70 kDa
DE protein) (HSP70).
GN DNAK.
OS Francisella tularensis.
OC Bacteria; Proteobacteria; gamma subdivision; Francisella group;
OC Francisella.
OX NCBI_TaxID=263;
RN SEQUENCE FROM N.A.
RX MEDLINE=96060855; PubMed=7590305;
RA Zuber M., Hoover T.A., Dartzbaugh M.T., Court D.L.;
RT "Analysis of the dnak molecular chaperone system of Francisella
RT tularensis.";
RL Gene 164:149-152(1995).
CC -1- FUNCTION: ACTS AS A CHAPERONE (BY SIMILARITY).
CC -1- INDUCTION: BY STRESS CONDITIONS E.G. HEAT SHOCK (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: LA3367; AAA69561.1; -.
CC HSSP: P04475; 1DKY.

```

```

DR PWMA-2DPAGE; P48205; -.
DR InterPro: IPR001023; Hsp70.
DR Pfam: PF00012; HSP70; 1.
DR PRINTS: PR00301; HEATSHOCK70.
DR ProDom: PD000089; HSP70; 1.
DR PROSITE: PS00297; HSP70_1; 1.
DR PROSITE: PS00329; HSP70_2; 1.
DR PROSITE: PS01036; HSP70_3; 1.
CC Chapterone; ATP-binding; Heat shock; Phosphorylation.
CC MOD RES 201 201 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
CC SEQUENCE 642 AA; 69208 MW; 7DCFA7B4714B96EA CRC64;
CC -----
Query Match 86.1%; Score 31; DB 1; Length 642;
Best Local Similarity 85.7%; Pred. No. 28;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 GQTRSP 7
DB 344 GQTRMPL 350
-----
RESULT 11
DNAX LEGPN STANDARD; PRT; 644 AA.
AC O32482;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Chapterone protein dnak (Heat shock protein 70) (Heat shock 70 kDa
DE protein) (HSP70).
GN DNAK.
OS Legionella pneumophila.
OC Bacteria; Proteobacteria; gamma subdivision; Legionellaceae group;
OC Legionellaceae; Legionella.
OX NCBI_TaxID=446;
RN SEQUENCE FROM N.A.
RP STRAIN=MS11;
RX MEDLINE=97473509; PubMed=9332363;
RA Amemura-Maekawa J., Watanabe H.;
RT "Cloning and sequencing of the dnak and grp genes of Legionella
RT pneumophila.";
RL Gene 197:165-168(1997).
CC -1- FUNCTION: ACTS AS A CHAPERONE (BY SIMILARITY).
CC -1- INDUCTION: BY STRESS CONDITIONS E.G. HEAT SHOCK (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D89498; BAA22783.1; -.
CC HSSP: P04475; 1DKY.
CC InterPro: IPR001023; Hsp70.
CC Pfam: PF00012; HSP70; 1.
CC PRINTS: PR00301; HEATSHOCK70.
CC ProDom: PD000089; HSP70; 1.
CC PROSITE: PS00297; HSP70_1; 1.
CC PROSITE: PS00329; HSP70_2; 1.
CC PROSITE: PS01036; HSP70_3; 1.
CC Chapterone; ATP-binding; Heat shock; Phosphorylation.
CC MOD RES 199 199 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
CC SEQUENCE 644 AA; 70087 MW; 8B3E748F09BEA757 CRC64;
CC -----
Query Match 86.1%; Score 31; DB 1; Length 644;
Best Local Similarity 85.7%; Pred. No. 28;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 GQTRSP 7

```

Db 342 GQTRMPL 348

# RESULT 12

AREA ASPOR STANDARD; PRT; 866 AA.

AC 013415;

DT 15-JUL-1998 (Rel. 36, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DE 15-OCT-2001 (Rel. 40, Last annotation update)

GN Nitrogen regulatory protein area.

OS Aspergillus oryzae.

OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;

OC Eurotiiales; Trichocommataceae; mitosporic Trichocommataceae; Aspergillus.

NCBI\_TaxID=5062;

RN

RP SEQUENCE FROM N.A.

RC STRAIN=IFO 4177;

RA Christensen T.; Hynes M.J.; Davis M.A.;

RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.

CC -1- FUNCTION: MAJOR NITROGEN REGULATORY PROTEIN. POSITIVELY ACTING

CC -1- SUBCELLULAR LOCATION: Nucleus.

CC -1- SIMILARITY: CONTAINS 1 GATA-TYPE ZINC FINGER.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@sb-sib.ch](mailto:license@sb-sib.ch)).

CC -----

DR EMBL; AJ002968; GAT05776.1; -

DR HSSP; P17429; 4GAT.

DR InterPro; IPR000679; Znf\_GATA.

DR Pfam; PF00320; GATA.1.

DR PRINTS; PR00619; GATAZNFINGER.

DR SMART; SM00401; ZNF\_GATA.1.

DR PROSITE; PS00344; GATA\_ZN\_FINGER\_1; 1.

DR PROSITE; PS50114; GATA\_ZN\_FINGER\_2; 1.

KM Transcription regulation; Activator; DNA-binding; Zinc-finger;

KM Nuclear protein; Nitrate assimilation.

FT ZN\_FING 664 688 GATA-TYPE

SO SEQUENCE 866 AA; 92882 MW; C09A03EB12B3FB84 CRC64;

Query Match 86.1%; Score 31; DB 1; Length 866;

Best Local Similarity 71.4%; Pred. No. 39;

Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

RESULT 13

PSMA\_HALNI STANDARD; PRT; 253 AA.

AC P57697;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DE 15-JUN-2002 (Rel. 41, Last annotation update)

DE Proteasome alpha subunit (EC 3.4.25.1) (Multicatalytic endopeptidase

DE complex alpha subunit).

GN PSMA OR VNG0166G.

OS Halobacterium sp. (strain NRC-1).

OC Archaea; Euryarchaeota; Halobacteriia; Halobacteriales;

OC Halobacteriaceae; Halobacterium.

NCBI\_TaxID=64091;

OX

RP SEQUENCE FROM N.A.

RL

RM

RS

RT

RU

RV

RW

RX

RY

RZ

RA

RB

RC

RD

RE

RF

RG

RH

RI

RJ

RK

RL

RM

RN

RO

RP

RS

RT

RU

RC STRAIN=K12 / MG1655;  
 RX MEDLINE=9534362; PubMed=7610040;  
 RA Burland V.D., Plunkett G. III, Sofia H.J., Daniels D.L.,  
 RA Blattner F.R.;  
 RT "Analysis of the Escherichia coli genome VI: DNA sequence of the  
 RT region from 92.8 through 100 minutes.";  
 RL Nucleic Acids Res. 23:2105-2119(1995).  
 RN [3]  
 RP FUNCTION.  
 RX MEDLINE=90060832; PubMed=2684786;  
 RA Webster C., Gardner L., Busby S.;  
 RT "The Escherichia coli *meiR* gene encodes a DNA-binding protein with  
 RT affinity for specific sequences located in the *mei*biose operon  
 RT regulatory region.";  
 RL Gene 83:207-213(1989).  
 RN [4]  
 RP MUTAGENESIS.  
 RX MEDLINE=93074985; PubMed=1445207;  
 RA Caswell R., Williams J., Lyddial A., Busby S.;  
 RT "Overexpression, purification and characterization of the Escherichia  
 RT coli *MeiR* transcription activator protein.";  
 RL Biochem. J. 287:493-499(1992).  
 RN [5]  
 RP CHARACTERIZATION.  
 RX MEDLINE=97263791; PubMed=9108148;  
 RA Bourgeois S.J., Michan C.M., Thomas M.S., Busby S.J.W., Hyde E.I.;  
 RT "DNA binding and DNA bending by the *MeiR* transcription activator  
 RT protein from Escherichia coli.";  
 RL Nucleic Acids Res. 25:1685-1693(1997).  
 CC -1- FUNCTION: TRANSCRIPTION ACTIVATOR FOR THE EXPRESSION OF THE MELAB  
 CC OPERON. *MEIR* BINDS AT TWO SITES LOCATED UPSTREAM OF THE MELAB  
 CC TRANSCRIPTION SITE.  
 CC POLARITY: BELONGS TO THE ARAC/XYLIS FAMILY OF TRANSCRIPTIONAL  
 CC REGULATORS.  
 CC CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: M18425; NOT ANNOTATED\_CDS.  
 CC EMBL: U14003; AAA97018.1; -.  
 CC EMBL: AE000464; AAC77079.1; -.  
 CC PIR: A29625; REGBMB.  
 CC DR ScGene; EG11230; *meiR*.  
 CC DR InterPro; IPR000005; HTHARAC.  
 CC DR Pfam; PF00165; HTH\_ARAC; 2.  
 CC DR PRINTS; PR00013; HTHARAC.  
 CC DR SMART; SM00342; HTH\_ARAC; 1.  
 CC DR PROSITE; PS00041; HTH\_ARAC\_FAMILY\_1; 1.  
 CC DR PROSITE; PS01124; HTH\_ARAC\_FAMILY\_2; 1.  
 CC KW Transcription regulation; DNA-binding; Activator; Complete proteome.  
 CC FT DNA BIND 210 229 H-T-H MOTIF (BY SIMILARITY).  
 CC FT MUTAGEN 222 222 N-XA: NO EFFECT.  
 CC FT MUTAGEN 269 269 R-XG: NO EFFECT.  
 CC FT MUTAGEN 271 271 S-XA: 50% REDUCED ACTIVITY.  
 CC FT MUTAGEN 272 272 S-XN: 75% REDUCED ACTIVITY.  
 CC SQ SEQUENCE 302 AA; 34928 MW; 2D9HC3069AB05029 CRC64;  
 CC  
 CC Query Match 83.3%; Score 30; DB 1; Length 302;  
 CC Best Local Similarity 100.0%; Pred. No. 21;  
 CC Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

ID CRB DROME STANDARD; PRT; 2139 AA.  
 AC P10040;  
 DT 01-MAR-1989 (Rel. 10, Created)  
 DT 01-MAY-1991 (Rel. 18, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Crumbs protein precursor (95F).  
 GN CRB.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;  
 OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
 OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Oregon-R; TISSUE=Embryo;  
 RX MEDLINE=90263104; PubMed=2344615;  
 RA Tepas U., Theres C., Knust E.;  
 RT "Crumbs encodes an EGF-like protein expressed on apical membranes of  
 RT Drosophila epithelial cells and required for organization of  
 RT epithelia.";  
 RL Cell 61:787-799(1990).  
 RN [2]  
 RP SEQUENCE OF 1663-1955 FROM N.A.  
 RC TISSUE=Embryo;  
 RX MEDLINE=87218537; PubMed=3107986;  
 RA Knust E., Dietrich U., Tepas U., Bremer K.A., Weigelt D.,  
 RA Vaesslin H., Campos-Ortega J.A.;  
 RT "EGF homologous sequences encoded in the genome of Drosophila  
 RT melanogaster, and their relation to neurogenic genes.";  
 RL EMBO J. 6:761-766(1987).  
 CC -1- FUNCTION: MAY PLAY A ROLE IN THE DEVELOPMENT OF EPITHELIA,  
 CC POSSIBLY FOR THE ESTABLISHMENT AND/OR MAINTENANCE OF CELL  
 CC POLARITY. IT MAY ACT AS A SIGNAL.  
 CC  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- PTM: PHOSPHORYLATED IN THE CYTOPLASMIC DOMAIN (POTENTIAL).  
 CC -1- SIMILARITY: CONTAINS 29 EGF-LIKE DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 3 LAMININ G-LIKE DOMAINS.  
 CC CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: M33753; AAA28428.1; ALT\_SEQ.  
 CC EMBL: X05144; CAA28793.1; -.  
 CC PIR: B26637; B26637.  
 CC PIR: A35672; A35672.  
 CC HSSP: P00740; 1EDM.  
 CC DR FlyBase; FBgn000368; *crb*.  
 CC DR InterPro; IPR000152; *Asx\_hydroxyl*.  
 CC DR InterPro; IPR000561; EGF\_1-like.  
 CC DR InterPro; IPR000742; EGF\_2.  
 CC DR InterPro; IPR001881; EGF\_Ca.  
 CC DR InterPro; IPR001438; EGF\_11.  
 CC DR InterPro; IPR001791; *Laminin\_G*.  
 CC DR Pfam; PF00054; *Laminin\_G*; 3.  
 CC DR PRINTS; PR00010; EGFALOOD.  
 CC DR Pfam; PF00008; EGF; 26.  
 CC DR SMART; SM00179; EGF\_CA; 11.  
 CC DR SMART; SM00001; EGF\_1like; 16.  
 CC DR SMART; SM00282; *LamG*; 3.  
 CC DR PROSITE; PS00010; ASX\_HYDROXYL; 15.  
 CC DR PROSITE; PS00022; EGF\_1; 26.  
 CC DR PROSITE; PS01186; EGF\_2; 17.  
 CC DR PROSITE; PS01187; EGF\_CA; 12.  
 CC DR PROSITE; PS00025; *LAM\_G\_DOMAIN*; 3.  
 CC KW Differentiation; Repeat; EGF-like domain; Transmembrane;  
 CC Glycoprotein; Signal; Phosphorylation.  
 CC FT SIGNAL 1 90  
 CC CHAIN 91 2139 CRUMBS PROTEIN.

RESULT 15  
 CRB\_DROME

```

FT DOMAIN 91 2084 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 2085 2111 POTENTIAL.
FT DOMAIN 2112 2139 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 267 303 EGF-LIKE 1.
FT DOMAIN 306 343 EGF-LIKE 2.
FT DOMAIN 348 386 EGF-LIKE 3.
FT DOMAIN 388 425 EGF-LIKE 4.
FT DOMAIN 427 463 EGF-LIKE 5.
FT DOMAIN 464 500 EGF-LIKE 6.
FT DOMAIN 501 532 EGF-LIKE 7.
FT DOMAIN 545 581 EGF-LIKE 8.
FT DOMAIN 582 611 EGF-LIKE 9.
FT DOMAIN 609 646 EGF-LIKE 10.
FT DOMAIN 648 685 EGF-LIKE 11.
FT DOMAIN 687 723 EGF-LIKE 12.
FT DOMAIN 725 761 EGF-LIKE 13.
FT DOMAIN 763 800 EGF-LIKE 14.
FT DOMAIN 802 838 EGF-LIKE 15.
FT DOMAIN 840 902 EGF-LIKE 16.
FT DOMAIN 904 940 EGF-LIKE 17.
FT DOMAIN 942 978 EGF-LIKE 18.
FT DOMAIN 980 1021 EGF-LIKE 19.
FT DOMAIN 1023 1205 LAMININ G-LIKE 1.
FT DOMAIN 1207 1243 EGF-LIKE 20.
FT DOMAIN 1250 1480 LAMININ G-LIKE 2.
FT DOMAIN 1481 1517 EGF-LIKE 21.
FT DOMAIN 1558 1758 LAMININ G-LIKE 3.
FT DOMAIN 1759 1795 EGF-LIKE 22.
FT DOMAIN 1797 1833 EGF-LIKE 23.
FT DOMAIN 1835 1871 EGF-LIKE 24.
FT DOMAIN 1874 1915 EGF-LIKE 25.
FT DOMAIN 1915 1951 EGF-LIKE 26.
FT DOMAIN 1953 1989 EGF-LIKE 27.
FT DOMAIN 1991 2029 EGF-LIKE 28.
FT DOMAIN 2030 2070 EGF-LIKE 29.
FT DOMAIN 271 282 BY SIMILARITY.
FT DISULFID 276 291 BY SIMILARITY.
FT DISULFID 293 302 BY SIMILARITY.
FT DISULFID 310 321 BY SIMILARITY.
FT DISULFID 315 331 BY SIMILARITY.
FT DISULFID 333 342 BY SIMILARITY.
FT DISULFID 352 363 BY SIMILARITY.
FT DISULFID 357 374 BY SIMILARITY.
FT DISULFID 376 385 BY SIMILARITY.
FT DISULFID 392 403 BY SIMILARITY.
FT DISULFID 397 412 BY SIMILARITY.
FT DISULFID 414 424 BY SIMILARITY.
FT DISULFID 431 442 BY SIMILARITY.
FT DISULFID 436 451 BY SIMILARITY.
FT DISULFID 453 462 BY SIMILARITY.
FT DISULFID 468 479 BY SIMILARITY.
FT DISULFID 473 488 BY SIMILARITY.
FT DISULFID 490 499 BY SIMILARITY.
FT DISULFID 505 515 BY SIMILARITY.
FT DISULFID 509 520 BY SIMILARITY.
FT DISULFID 522 531 BY SIMILARITY.
FT DISULFID 549 562 BY SIMILARITY.
FT DISULFID 556 569 BY SIMILARITY.
FT DISULFID 571 580 BY SIMILARITY.
FT DISULFID 586 597 BY SIMILARITY.
FT DISULFID 591 602 BY SIMILARITY.
FT DISULFID 604 610 BY SIMILARITY.
FT DISULFID 613 624 BY SIMILARITY.
FT DISULFID 618 634 BY SIMILARITY.
FT DISULFID 636 645 BY SIMILARITY.
FT DISULFID 652 664 BY SIMILARITY.
FT DISULFID 659 673 BY SIMILARITY.
FT DISULFID 675 684 BY SIMILARITY.
FT DISULFID 691 702 BY SIMILARITY.
FT DISULFID 702 711 BY SIMILARITY.
FT DISULFID 713 722 BY SIMILARITY.
FT DISULFID 729 740 BY SIMILARITY.
FT DISULFID 734 749 BY SIMILARITY.

```

```

FT DISULFID 751 760 BY SIMILARITY.
FT DISULFID 767 778 BY SIMILARITY.
FT DISULFID 772 787 BY SIMILARITY.
FT DISULFID 789 799 BY SIMILARITY.
FT DISULFID 806 817 BY SIMILARITY.
FT DISULFID 811 826 BY SIMILARITY.
FT DISULFID 828 837 BY SIMILARITY.
FT DISULFID 844 855 BY SIMILARITY.
FT DISULFID 849 890 BY SIMILARITY.
FT DISULFID 892 901 BY SIMILARITY.
FT DISULFID 908 919 BY SIMILARITY.
FT DISULFID 913 928 BY SIMILARITY.
FT DISULFID 930 939 BY SIMILARITY.
FT DISULFID 946 957 BY SIMILARITY.
FT DISULFID 968 977 BY SIMILARITY.
FT DISULFID 984 995 BY SIMILARITY.
FT DISULFID 989 1009 BY SIMILARITY.
FT DISULFID 1011 1020 BY SIMILARITY.
FT DISULFID 1211 1222 BY SIMILARITY.
FT DISULFID 1216 1231 BY SIMILARITY.
FT DISULFID 1233 1242 BY SIMILARITY.
FT DISULFID 1285 1294 BY SIMILARITY.
FT DISULFID 1485 1496 BY SIMILARITY.
FT DISULFID 1490 1505 BY SIMILARITY.
FT DISULFID 1507 1516 BY SIMILARITY.
FT DISULFID 1763 1774 BY SIMILARITY.
FT DISULFID 1768 1783 BY SIMILARITY.
FT DISULFID 1785 1794 BY SIMILARITY.
FT DISULFID 1801 1812 BY SIMILARITY.
FT DISULFID 1806 1821 BY SIMILARITY.
FT DISULFID 1823 1832 BY SIMILARITY.
FT DISULFID 1839 1850 BY SIMILARITY.
FT DISULFID 1844 1859 BY SIMILARITY.
FT DISULFID 1861 1870 BY SIMILARITY.
FT DISULFID 1878 1889 BY SIMILARITY.
FT DISULFID 1883 1903 BY SIMILARITY.
FT DISULFID 1905 1914 BY SIMILARITY.
FT DISULFID 1919 1930 BY SIMILARITY.
FT DISULFID 1924 1939 BY SIMILARITY.
FT DISULFID 1941 1950 BY SIMILARITY.
FT DISULFID 1957 1968 BY SIMILARITY.
FT DISULFID 1962 1977 BY SIMILARITY.
FT DISULFID 1979 1988 BY SIMILARITY.
FT DISULFID 1995 2008 BY SIMILARITY.
FT DISULFID 2002 2017 BY SIMILARITY.
FT DISULFID 2019 2028 BY SIMILARITY.
FT CARBOHYD 37 37 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 96 96 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 198 198 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 238 238 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 239 239 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 336 336 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 400 400 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 550 550 N-LINKED (GLCNAC. .) (POTENTIAL).

```

```

Query Match 83.3%; Score 30; DB 1; Length 2139;
Best Local Similarity 85.7%; Pred. No. 1.8e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 1 GQRRSPL 7
DB 1551 GQRRSPL 1557

```

```

Search completed: January 29, 2003, 14:00:26
Job time : 5.84615 secs

```

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2003, 13:56:21 ; Search time 19.5641 seconds  
(without alignments)  
73.723 Million cell updates/sec

Title: US-09-807-949a-109  
Perfect score: 36  
Sequence: 1 GQTRSP 7

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

SPTREMBL\_21:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	34	94.4	216	14	P97969
2	33	91.7	949	5	Q9Y0H4
3	32	88.9	144	17	Q9HRN7
4	32	88.9	159	2	Q9WKC6
5	32	88.9	311	12	O8V719
6	32	88.9	401	6	Q9SKT9
7	32	88.9	425	4	Q9NKE8
8	32	88.9	472	5	Q16804
9	32	88.9	844	5	O16117
10	31	86.1	155	5	O18010
11	31	86.1	164	2	O8RTK8
12	31	86.1	168	2	Q8RTM9
13	31	86.1	221	14	P97971
14	31	86.1	304	4	Q8TDK0
15	31	86.1	422	15	Q8URT3
16	31	86.1	423	15	O12272

17	31	86.1	647	2	Q9L515	Q9L515 psychobact
18	31	86.1	683	13	Q9UA35	Q9UA35 eptaretus
19	31	86.1	691	4	Q9H831	Q9H831 homo sapien
20	31	86.1	703	11	Q9ERH8	Q9ERH8 mus musculu
21	31	86.1	703	11	Q9IVD7	Q9IVD7 mus musculu
22	31	86.1	705	11	Q8VIR3	Q8VIR3 ratu musc
23	31	86.1	853	3	Q9C414	Q9C414 aspergillus
24	31	86.1	866	3	Q9Y7E8	Q9Y7E8 aspergillus
25	31	86.1	867	12	Q91RA9	Q91RA9 bovine herp
26	30	83.3	106	2	Q51656	Q51656 paracoccus
27	30	83.3	246	16	Q9XIC6	Q9XIC6 thermocoga
28	30	83.3	270	16	Q9KAR9	Q9KAR9 bacillus ha
29	30	83.3	302	16	Q8XD02	Q8XD02 escherichia
30	30	83.3	310	2	Q9F1K3	Q9F1K3 citrobacter
31	30	83.3	310	16	Q8XH24	Q8XH24 salmonella
32	30	83.3	331	16	Q8UR92	Q8UR92 agrobacteri
33	30	83.3	405	16	Q9PCN8	Q9PCN8 xylella fas
34	30	83.3	408	16	Q9A0U4	Q9A0U4 streptococc
35	30	83.3	478	5	Q9Y1X5	Q9Y1X5 ephydalia f
36	30	83.3	518	5	O16283	O16283 caenorhabdi
37	30	83.3	602	5	O46138	O46138 lineus sang
38	30	83.3	665	16	Q93GP5	Q93GP5 salmonella
39	30	83.3	670	5	Q9W474	Q9W474 drosophila
40	30	83.3	899	16	P74477	P74477 synechocyst
41	30	83.3	1013	12	Q91LX9	Q91LX9 retroperito
42	30	83.3	1385	5	Q8WT26	Q8WT26 leishmania
43	30	83.3	1703	3	Q8TG39	Q8TG39 exophiala d
44	30	83.3	1806	16	Q929J3	Q929J3 listeria in
45	30	83.3	2146	5	Q9VC97	Q9VC97 drosophila

#### ALIGNMENTS

RESULT 1  
ID P97969 PRELIMINARY, PRT, 216 AA.  
AC P97969;  
DT 01-MAY-1997 (TREMBLrel. 03, Created)  
DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)  
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
DE 70-kDa heat shock protein (Fragment).  
GN HSP70.  
OS unidentified soil organism.  
OC unclassified; environmental samples.  
OX NCB1\_Taxid=46465;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Yap W.H., Li X., Soong T.W., Davies J.E.;  
RT \*Genetic diversity of soil microorganisms assessed by analysis of  
RT hept0 (hna) sequences.";  
RL Submitted (FBI-1996) to the EMBL/GenBank/DBJ databases.  
DR EMBL, U49138; AAB48226.1; -.  
DR InterPro, IPR01023; Hsp70.  
DR Pfam, PF00012; HSP70; 1.  
DR ProDom, PD000089; Hsp70; 1.  
DR PROSITE, PS00329; HSP70\_2; 1.  
KW Heat shock.  
FT NON\_TER 1  
FT NON\_TER 216  
SQ SEQUENCE 216 AA; 23717 MW; F3AF5E1A6AB4EC36 CRC64;  
Query Match Score 34; DB 14; Length 216;  
Best Local Similarity 85.7%; Pred. No. 8;  
Matches 6; Conservative 1; Mismatches 0; Indels 0;  
QY 1 GQTRSP 7  
DB 186 GQTRSP 192

RESULT 2  
Q9Y0H4

ID Q9Y0H4 PRELIMINARY; PRT; 949 AA.  
 AC Q9Y0H4;  
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Su(dx) protein.  
 GN SU(DX) OR CG4244.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OC NCBI\_TaxId=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Mortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Aroll J.F., Agbayani A., An H.-U., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Baas A., Baxendale J., Bayraktaroglu I., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolintsov S.,  
 RA Borkova D., Botchan D.A., Butler H., Cadden E., Center A., Chandra I.,  
 RA Burtis K.C., Buam D.A., Butler H., Cadden E., Center A., Chandra I.,  
 RA Cherry J.M., Caley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Dou P.L.B., Downes M., Dugan-Rocha S., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Fowler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jaitai M., Kalush F., Karpen G.H., Ke Z., Kemison J.A., Kerchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Paclob J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier B., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svatek R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasmann D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers R.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster.";  
 RL Science 287:2185-2195(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Cornell M., Evans D.A.P., Mann R., Foster M., Flaszka M.,  
 RA Montchacon M., Artavanis-Tsakonas S., Baron M.;  
 RT "The Drosophila melanogaster Suppressor of deltex gene, a regulator of  
 RT the Notch receptor signalling pathway, is an E3 class ubiquitin  
 RT ligase.";  
 RL Genetics 152:0-0(1999).  
 RT EMBL: AE003584; AAF51312.1; -  
 DR EMBL: AFI52865; AAD38975.1; -  
 DR HSSP: Q13526; IPIN.  
 DR PLYBase: FBgn0003557; Su(dx).  
 DR InterPro: IPR000008; C2.  
 DR InterPro: IPR000569; HECT domain.  
 DR InterPro: IPR001202; WW\_Reps\_WMP.  
 DR Pfam: PF00166; C2; 1.  
 DR Pfam: PF00632; HECT; 1.  
 DR Pfam: PF00397; WW; 4.  
 DR SMART: SMO0239; C2; 1.

DR SMART: SMO0119; HECTC; 1.  
 DR SMART: SMO0456; WW; 3.  
 DR PROSITE: PSS0004; C2\_DOMAIN\_2; 2.  
 DR PROSITE: PSS0237; HECT; 2.  
 DR PROSITE: PSS0159; WW\_DOMAIN\_1; 3.  
 DR PROSITE: PSS0020; WW\_DOMAIN\_2; 4.  
 SQ SEQUENCE 949 AA; 107966 MW; 74817A8B05AC6E6B CRC64;  
 Query Match 91.7%; Score 33; DB 5; Length 949;  
 Best Local Similarity 85.7%; Pred. No. 63;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GQTRSP 7  
 DB 236 GQTRSP 242  
 RESULT 3  
 ID Q9HRN7 PRELIMINARY; PRT; 144 AA.  
 AC Q9HRN7;  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)  
 DE Vng00613h.  
 GN VNG00613h.  
 OS Halobacterium sp. (strain NRC-1).  
 OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;  
 OC Halobacterium.  
 OC NCBI\_TaxId=64091;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20504483; PubMed=11016950;  
 RA Ng W.V., Kennedy S.P., Mahaitas G.G., Berquist B., Pan M.,  
 RA Shukla H.D., Laskey S.R., Baliga N.S., Thorsen V., Shroga J.,  
 RA Sartzell B., Weir D., Hall J., Dahl T.A., Weir R., Goo Y.A.,  
 RA Leitzauer B., Keller K., Cruz R., Danon M.J., Hough D.W.,  
 RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Jung K.-H.,  
 RA Tembarger T.A., Beck R.F., Pohlschroder M., Spudich J.L., Dale H.,  
 RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,  
 RA Edhardt H., Lowe T.M., Liang P., Riley M., Hood L., Dasatma S.;  
 RT "Genome sequence of Halobacterium species NRC-1.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).  
 DR EMBL: AE005010; AAG19121.1; -  
 KW Complete proteome.  
 SQ SEQUENCE 144 AA; 15500 MW; 72B8A3D7CC9B2AB CRC64;  
 Query Match 88.9%; Score 32; DB 17; Length 144;  
 Best Local Similarity 85.7%; Pred. No. 15;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 GQTRSP 7  
 DB 134 GQTRSP 140  
 RESULT 4  
 ID Q9MXC6 PRELIMINARY; PRT; 159 AA.  
 AC Q9MXC6;  
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE WctL-like protein.  
 GN ORF1.  
 OS Pseudomonas hydrogencitromphila.  
 OC Bacteria; Proteobacteria; beta subdivision; Hydrogenophilus group;  
 OC Hydrogenophilus.  
 OC NCBI\_TaxId=297;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=TH-1;  
 RA Hayashi N.R., Terazono K., Yokoyama K., Kodama T., Igatahshi Y.;



RT "Analysis of the genes located upstream and downstream of Rubisco  
RT genes in thermophilic hydrogen-oxidizing bacterium, Hydrogenophilus  
RT thermoluteus.";  
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.

DR EMBL; AB018441; BAA76604.1; -;  
DR InterPro; IPR000086; NUDIX\_hydrolase.

DR Pfam; PF00293; NUDIX; 1.  
DR PROSITE; PS00893; NUDIX; 1.

SO SEQUENCE 159 AA; 17725 MW; 080728CB05DDC796 CRC64;

Query Match 88.9%; Score 32; DB 2; Length 159;  
Best Local Similarity 85.7%; Pred. No. 17;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GQTRSP 7  
Db 123 GQTRSP 129

#### RESULT 5

ID 08V719 PRELIMINARY; PRT; 311 AA.  
AC 08V719;  
DT 01-MAR-2002 (TEMBLrel. 20, Created)  
DT 01-MAR-2002 (TEMBLrel. 20, Last sequence update)  
DT 01-MAR-2002 (TEMBLrel. 20, Last annotation update)  
DE Tegument protein.  
GN US10.  
OS Simian herpes B virus (Cercopithecid herpesvirus 1) (Shbv).  
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
OC Alphaherpesvirinae; Simplexvirinae.  
OX NCBI\_TaxID=10325;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=E2490;  
RA MEDLINE=21635528; PubMed=11773425;  
RA Ohsawa K., Black D.H., Sato H., Eberle R.;  
RT "Sequence and Genetic Arrangement of the Us Region of the Monkey B  
RT Virus (Cercopithecine herpesvirus 1) Genome and Comparison with the Us  
RT Region of Other Primate Herpesviruses.";  
RL J. Virol. 76:1516-1520 (2002).  
DR EMBL; AB074432; BAB83758.1; -;  
SO SEQUENCE 311 AA; 34034 MW; B407F49FB4B47FB CRC64;

Query Match 88.9%; Score 32; DB 12; Length 311;  
Best Local Similarity 100.0%; Pred. No. 34;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GQTRSP 6  
Db 102 GQTRSP 107

#### RESULT 6

ID 095KT9 PRELIMINARY; PRT; 401 AA.  
AC 095KT9;  
DT 01-DEC-2001 (TEMBLrel. 19, Created)  
DT 01-DEC-2001 (TEMBLrel. 19, Last sequence update)  
DT 01-MAR-2002 (TEMBLrel. 20, Last annotation update)  
DE Hypothetical 44.6 kDa protein.  
OS Macaca fascicularis (creb eating macaque) (Cynomolgus monkey).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;  
OC Cercopithecinae; Macaca.  
OX NCBI\_TaxID=9541;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=TEMPORAL LOBE RIGHT;  
RA Osada N., Hida M., Kuseida J., Tanuma R., Iseki K., Hirai M., Terao K.,  
RA Suzuki Y., Sugano S., Hashimoto K.;  
RT "Isolation of full-length cDNA clones from macaque brain cDNA  
RT libraries.";

RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AB060844; BAB46867.1; -;  
DR InterPro; IPR001837; CAP.  
DR InterPro; IPR001670; Fe-ADH.  
DR InterPro; IPR000215; Serpin.  
DR Pfam; PF01213; CAP; 1.  
DR PROSITE; PS00060; ADH\_IRON\_2; UNKNOWN\_1.  
DR PROSITE; PS01089; CAP\_2; UNKNOWN\_1.  
DR PROSITE; PS00284; SERPIN; UNKNOWN\_1.  
KM Hypothetical protein.  
SO SEQUENCE 401 AA; 44620 MW; 6F57F071FF695387 CRC64;

Query Match 88.9%; Score 32; DB 6; Length 401;  
Best Local Similarity 100.0%; Pred. No. 44;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GQTRSP 6  
Db 221 GQTRSP 226

#### RESULT 7

ID 09NXE8 PRELIMINARY; PRT; 425 AA.  
AC 09NXE8;  
DT 01-OCT-2000 (TEMBLrel. 15, Created)  
DT 01-OCT-2000 (TEMBLrel. 15, Last sequence update)  
DT 01-MAR-2002 (TEMBLrel. 20, Last annotation update)  
DE cDNA FLJ20291 fls, clone HEP04788 (Hypothetical 49.6 kDa  
DE protein).  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T., Matsumura K.,  
RA Nakajima Y., Mizuno T., Morinaga M., Ota T., Suzuki M.,  
RA Nishi T., Shibahara T., Tanaka T., Nakamura Y., Isegaki T., Sugano S.;  
RT "WEDD human cDNA sequencing project.";  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=UTERUS;  
RA Strausberg R.;  
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AK000298; BAA91065.1; -;  
DR EMBL; BC008833; AA088833.1; -;  
KM Hypothetical protein.  
SO SEQUENCE 425 AA; 49647 MW; F8288E143A3FF764 CRC64;

Query Match 88.9%; Score 32; DB 4; Length 425;  
Best Local Similarity 100.0%; Pred. No. 47;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GQTRSP 6  
Db 312 GQTRSP 317

#### RESULT 8

ID 016804 PRELIMINARY; PRT; 472 AA.  
AC 016804;  
DT 01-JAN-1998 (TEMBLrel. 05, Created)  
DT 01-JAN-1998 (TEMBLrel. 05, Last sequence update)  
DT 01-MAR-2002 (TEMBLrel. 20, Last annotation update)  
DE Paired box protein (fragment).  
GN SV OR DPX258 OR CG1049.  
OS Drosophila melanogaster (fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.

```

OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Cherry T., Bouchard M., Kozmik Z., Busslinger M.;
RL Mech. Dev. 0:0-0(1997).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -1- SIMILARITY: CONTAINS A PAIRED BOX DOMAIN.
DR EMBL: AF016888; AAB70249.1; -.
DR HSSP: P26367; 6PAX.
DR FlyBase; FBgn0005561; sv.
DR InterPro; IPR001523; Paired_box.
DR Pfam; PF00292; PAX; 1.
DR PRINTS; PR00027; PAIREDBOX.
DR SMART; SM00351; PAX; 1.
DR PROSITE; PS00034; PAIRED BOX; 1.
DR DMB-binding; Developmental protein; Nuclear protein; Paired box;
KM Transcription regulation.
FT NON_TER
SQ SEQUENCE 472 AA; 50043 MW; C02676C41D981D8D CRC64;

Query Match 88.9%; Score 32; DB 5; Length 472;
Best Local Similarity 85.7%; Pred. No. 52;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GQTRSPL 7
Db 427 GQTSPL 433

RESULT 9
ID 016117 PRELIMINARY; PRT; 844 AA.
AC 016117;
DT 01-JAN-1998 (TRENBLrel. 05, Created)
DT 01-JAN-1998 (TRENBLrel. 05, Last sequence update)
DT 01-MAR-2002 (TRENBLrel. 20, Last annotation update)
DE sv gene product.
GN SV OR SPA OR CG11049.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Mortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champagne M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abdi J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bertan B.P., Bhandari D., Bolshakov S.,
RA Bokora D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Dou L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira W., Fleischmann W.,
RA Foster C., Gabrielian A.B., Garg N.S., Gelbart W.M., Glasner K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heitman T.J., Hernandez J.R., Houck J.,
RA Houten R., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
RA Jaisi M., Kalush F., Karpen G.H., Ke Z., Kemnison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Pauleb J.M.,

```

```

RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier B., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svrtkars R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weisenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=97427858; PubMed=9284046;
RA Fu W., Noll M.;
RT "The Pax2 homolog sparkling is required for development of cone and
RT pigment cells in the Drosophila eye.";
RL Genes Dev. 11:2066-2078(1997).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -1- SIMILARITY: CONTAINS A PAIRED BOX DOMAIN.
DR EMBL: AF010256; AAB86598.1; -.
DR EMBL: AF010256; AAB86598.1; -.
DR HSSP: P26367; 6PAX.
DR FlyBase; FBgn0005561; sv.
DR InterPro; IPR001523; Paired box.
DR InterPro; IPR00130; Zn_MTpeptidase.
DR Pfam; PF00292; PAX; 1.
DR PRINTS; PR00027; PAIREDBOX.
DR SMART; SM00351; PAX; 1.
DR PROSITE; PS00034; PAIRED BOX; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN 1.
KM DNA-binding; Developmental protein; Nuclear protein; Paired box;
KM Transcription regulation.
SQ SEQUENCE 844 AA; 90764 MW; B4B4EBB424A21204 CRC64;

Query Match 88.9%; Score 32; DB 5; Length 844;
Best Local Similarity 85.7%; Pred. No. 95;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GQTRSPL 7
Db 799 GQTSPL 805

RESULT 10
ID 018010 PRELIMINARY; PRT; 155 AA.
AC 018010;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE Hypothetical 17.9 kDa protein.
GN C15C7.4.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Pelodetidae; Caenorhabditis.
OX NCBI_TaxID=6239;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Leimbach D.;
RT "The sequence of C. elegans cosmid C15C7.";
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
[3]

```

RP SEQUENCE FROM N.A.  
RC STRAIN=BRISTOL N2;  
RA Watson R.;  
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; U41528; AAK39166.1; -.  
KW Hypothetical protein.  
SQ SEQUENCE 155 AA; 17947 MW; 280455D1101A838E CRC64;

Query Match 86.1%; Score 31; DB 5; Length 155;  
Best Local Similarity 85.7%; Pred. No. 28;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GQTRSP 7  
Db 59 GQTRCPL 65

RESULT 11  
Q8RJK8 PRELIMINARY; PRT; 164 AA.

AC Q8RJK8;  
DT 01-JUN-2002 (T-EMBLrel. 21, Created)  
DT 01-JUN-2002 (T-EMBLrel. 21, Last sequence update)  
DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)  
DE Wave (Fragment).  
GN WAVE.  
OS Vibrio cholerae.  
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.  
ON NCBI\_TaxID=666;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN=V194;  
RA Nesper J., Kraiss A., Schild S., Blass J., Klose K.E., Bockemuhl J.,  
RT "Comparative and genetic analysis of the putative Vibrio cholerae LPS  
core oligosaccharide biosynthesis (wav) gene cluster."  
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF444793; AAL77355.2; -.  
FT NON TER 1  
SQ SEQUENCE 164 AA; 19418 MW; 63DA9F9B37357E77 CRC64;

Query Match 86.1%; Score 31; DB 2; Length 164;  
Best Local Similarity 85.7%; Pred. No. 29;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GQTRSP 7  
Db 136 GQTRRPL 142

RESULT 12  
Q8RJM9 PRELIMINARY; PRT; 168 AA.

AC Q8RJM9;  
DT 01-JUN-2002 (T-EMBLrel. 21, Created)  
DT 01-JUN-2002 (T-EMBLrel. 21, Last sequence update)  
DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)  
DE Wave (Fragment).  
GN WAVE.  
OS Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.  
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.  
ON NCBI\_TaxID=666;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN=V215;  
RA Nesper J., Kraiss A., Schild S., Blass J., Klose K.E., Bockemuhl J.,  
RT "Comparative and genetic analysis of the putative Vibrio cholerae LPS  
core oligosaccharide biosynthesis (wav) gene cluster."  
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF443426; AAL76929.1; -.  
FT NON TER 1  
SQ SEQUENCE 168 AA; 19968 MW; 48A3A1A3F2FA2BB4 CRC64;

Query Match 86.1%; Score 31; DB 2; Length 168;  
Best Local Similarity 85.7%; Pred. No. 30;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GQTRSP 7  
Db 140 GQTRRPL 146

RESULT 13  
P97971 PRELIMINARY; PRT; 221 AA.

AC P97971;  
DT 01-MAY-1997 (T-EMBLrel. 03, Created)  
DT 01-MAY-1997 (T-EMBLrel. 03, Last sequence update)  
DT 01-MAY-2002 (T-EMBLrel. 20, Last annotation update)  
DE 70-kDa heat shock protein (Fragment).  
GN HSP70.  
OS unidentified soil organism.  
OC unclassified; environmental samples.  
ON NCBI\_TaxID=46465;  
RN (1)  
RP SEQUENCE FROM N.A.  
RA Yap W.H., Li X., Soong T.W., Davies J.E.;  
RT "Genetic diversity of soil microorganisms assessed by analysis of  
hap70 (dnaK) sequences."  
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.  
DR EMBL; U49140; AAB48228.1; -.  
DR InterPro; IPR01023; HSP70.  
DR Pfam; PF00012; HSP70; 1.  
DR ProDom; PD000089; HSP70; 1.  
DR PROSITE; PS00329; HSP70\_2; 1.  
DR PROSITE; PS01036; HSP70\_3; 1.  
KW Heat shock.  
FT NON TER 1  
FT NON TER 221  
SQ SEQUENCE 221 AA; 23905 MW; 901CBCE9B29769F3 CRC64;

Query Match 86.1%; Score 31; DB 14; Length 221;  
Best Local Similarity 85.7%; Pred. No. 40;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GQTRSP 7  
Db 191 GQTRMPL 197

RESULT 14  
Q8TDKO PRELIMINARY; PRT; 304 AA.

AC Q8TDKO;  
DT 01-JUN-2002 (T-EMBLrel. 21, Created)  
DT 01-JUN-2002 (T-EMBLrel. 21, Last sequence update)  
DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)  
DE Gml48 form A. (Human).  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
ON NCBI\_TaxID=9606;  
RN (1)  
RP SEQUENCE FROM N.A.  
RX MEDLINE=2188635; PubMed=11891061;  
RA Makalowska I., Sood R., Faruque M.U., Hu P., Robb C.M.,  
RT "Identification of six novel genes by experimental validation of  
RT Gene machine predicted genes."  
RL Gene 284:203-213(2002).  
DR EMBL; AF387617; AAL91358.1; -.  
SQ SEQUENCE 304 AA; 32908 MW; 27D17322971ED874 CRC64;

Query Match 86.1%; Score 31; DB 4; Length 304;  
Best Local Similarity 85.7%; Pred. No. 56;

Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GOTRSPL 7

Db 265 GOTRSPL 271

## RESULT 15

OBURT3  
ID OBURT3 PRELIMINARY; PRT; 422 AA.  
AC OBURT3;  
DT 01-MAR-2002 (TRENBLrel. 20, Created)  
DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)  
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)  
DE Gag polypeptide.  
GN GAG.  
OS Primate T-lymphotropic virus 3.  
OC Viruses; Retrovirdae; Deltaretrovirus.  
OX NCBI\_TaxID=194443;  
RN [1]  
RP SEQUENCE FROM N. A.  
RC STRAIN=CTO-604;  
RX MEDLINE=21602556; PubMed=11739691;  
RA Meertens L., Mahieux R., Mauciere P., Lewis J., Gessain A.  
RT "Complete Sequence of a Novel Highly Divergent Simian T-Cell  
RT Lymphotropic Virus from Wild-Caught Red-Capped Mangabeys (*Cercopithecus  
RT torquatus*) from Cameroon: a New Primate T-Lymphotropic Virus Type 3  
RT Subtype."  
RL J. Virol. 76:259-268 (2002).  
DR EMBL; AF391797; AAL48211.1;  
DR InterPro; IPR003139; Gag\_P19.  
DR InterPro; IPR000721; Gag\_P24.  
DR InterPro; IPR01878; Znf\_CCHC.  
DR Pfam; PF02228; Gag\_P19; 1.  
DR Pfam; PF00607; Gag\_P24; 1.  
DR Pfam; PF00098; Zf\_CCHC; 2.  
DR PRINTS; PR00939; C2HCZNFINGER.  
DR SMART; SM00343; Znf\_C2HC; 2.  
DR PROSITE; PS50158; ZF\_CCHC; 1.  
KW Polyprotein.  
SQ SEQUENCE 422 AA; 46868 MW; D0359F2D8ACE58DF CRC64;

Query Match 86.1%; Score 31; DB 15; Length 422;  
Best Local Similarity 85.7%; Pred. No. 78;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GOTRSPL 7

Db 312 GOTNSPL 318

Search completed: January 29, 2003, 14:02:24  
Job time : 20.5641 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2003, 13:57:26 ; Search time 8.61539 Seconds  
(without alignments)  
23.906 Million cell updates/sec

Title: US-09-807-949a-109

Perfect score: 36

Sequence: 1 GQTRSP 7

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 262574 seqs, 29422922 residues

262574

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents, AA:\*  
1: /cgn2\_6/prodata/1/1aa/5A\_COMB.pep:\*  
2: /cgn2\_6/prodata/1/1aa/5B\_COMB.pep:\*  
3: /cgn2\_6/prodata/1/1aa/6A\_COMB.pep:\*  
4: /cgn2\_6/prodata/1/1aa/6B\_COMB.pep:\*  
5: /cgn2\_6/prodata/1/1aa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/prodata/1/1aa/backfile1.pep:\*

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	36	100.0	7	US-09-178-115-109	Sequence 109, App
2	36	100.0	7	US-09-177-776-109	Sequence 109, App
3	31	86.1	853	US-09-254-325-2	Sequence 2, Appl
4	31	86.1	866	US-09-079-415-6	Sequence 6, Appl
5	31	86.1	866	US-08-750-458A-2	Sequence 2, Appl
6	30	83.3	7	US-09-178-115-107	Sequence 107, App
7	30	83.3	7	US-09-177-776-107	Sequence 107, App
8	30	83.3	7	US-09-084-605B-29	Sequence 29, Appl
9	30	83.3	7	US-09-438-150-6	Sequence 6, Appl
10	30	83.3	295	US-09-134-001C-2850	Sequence 2850, Ap
11	29	80.6	7	US-09-178-115-108	Sequence 108, App
12	29	80.6	7	US-09-177-776-108	Sequence 108, App
13	29	80.6	185	US-08-044-621D-36	Sequence 36, Appl
14	29	80.6	185	US-08-044-621D-37	Sequence 37, Appl
15	29	80.6	185	US-08-709-912-3	Sequence 3, Appl
16	29	80.6	185	US-08-709-912-5	Sequence 5, Appl
17	29	80.6	185	US-09-047-370-3	Sequence 3, Appl
18	29	80.6	185	US-09-047-370-5	Sequence 5, Appl
19	29	80.6	208	US-08-315-695-21	Sequence 21, Appl
20	29	80.6	213	US-08-104-445-3	Sequence 3, Appl
21	29	80.6	278	US-09-260-283-2	Sequence 2, Appl
22	29	80.6	626	US-09-357-251-34	Sequence 34, Appl
23	29	80.6	1536	US-09-413-814-10	Sequence 10, Appl
24	28	77.8	12	US-08-471-780C-35	Sequence 35, Appl
25	28	77.8	12	US-08-467-282B-35	Sequence 35, Appl
26	28	77.8	12	US-08-471-282A-35	Sequence 35, Appl
27	28	77.8	12	US-08-466-710C-35	Sequence 35, Appl

28	28	77.8	12	3	US-08-468-739C-35	Sequence 35, Appl
29	28	77.8	18	1	US-08-471-780C-36	Sequence 36, Appl
30	28	77.8	18	1	US-08-467-282B-36	Sequence 36, Appl
31	28	77.8	18	2	US-08-471-282A-36	Sequence 36, Appl
32	28	77.8	18	2	US-08-466-710C-36	Sequence 36, Appl
33	28	77.8	18	3	US-08-468-739C-36	Sequence 36, Appl
34	28	77.8	27	1	US-08-245-853-10	Sequence 10, Appl
35	28	77.8	27	1	US-08-573-675-10	Sequence 10, Appl
36	28	77.8	43	1	US-08-471-780C-54	Sequence 54, Appl
37	28	77.8	43	1	US-08-467-282B-54	Sequence 54, Appl
38	28	77.8	43	2	US-08-471-282A-54	Sequence 54, Appl
39	28	77.8	43	2	US-08-468-710C-54	Sequence 54, Appl
40	28	77.8	43	3	US-08-468-739C-54	Sequence 54, Appl
41	28	77.8	137	4	US-09-319-056B-29	Sequence 29, Appl
42	27	75.0	247	4	US-09-370-838-112	Sequence 112, Appl
43	27	75.0	325	4	US-09-041-886-33	Sequence 33, Appl
44	27	75.0	452	4	US-09-134-001C-4173	Sequence 4173, Ap
45	27	75.0	635	1	US-07-832-855-2	Sequence 2, Appl

## ALIGNMENTS

RESULT 1  
US-09-178-115-109  
Sequence 109, Application US/09178115

Patent No. 6297041

GENERAL INFORMATION:

APPLICANT: Zavada, Jan

APPLICANT: Pastorekova, Silvia

TITLE OF INVENTION: MN Gene and Protein

FILE REFERENCE: D-0021.5A

CURRENT APPLICATION NUMBER: US/09/178,115

EARLIER FILING DATE: 1998-10-23

EARLIER APPLICATION NUMBER: 09/177,776

EARLIER FILING DATE: 1998-10-23

EARLIER APPLICATION NUMBER: 08/787,739

EARLIER FILING DATE: 1997-01-24

EARLIER APPLICATION NUMBER: 08/485,049

EARLIER FILING DATE: 1995-06-07

EARLIER APPLICATION NUMBER: 08/486,756

EARLIER FILING DATE: 1995-06-07

EARLIER APPLICATION NUMBER: 08/477,504

EARLIER FILING DATE: 1995-06-07

EARLIER APPLICATION NUMBER: 08/481,658

EARLIER FILING DATE: 1995-06-07

EARLIER APPLICATION NUMBER: 08/485,862

EARLIER FILING DATE: 1995-06-07

EARLIER APPLICATION NUMBER: 08/485,863

EARLIER FILING DATE: 1995-06-07

EARLIER APPLICATION NUMBER: 08/487,077

EARLIER FILING DATE: 1995-06-07

EARLIER APPLICATION NUMBER: 08/260,190

EARLIER FILING DATE: 1994-06-15

EARLIER APPLICATION NUMBER: 08/177,093

EARLIER FILING DATE: 1993-12-30

EARLIER APPLICATION NUMBER: 07/964,589

EARLIER FILING DATE: 1992-10-21

EARLIER APPLICATION NUMBER: PV-709-92

EARLIER FILING DATE: 1992-03-11

NUMBER OF SEQ ID NOS: 116

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 109

LENGTH: 7

TYPE: PRT

ORGANISM: HUMAN

US-09-178-115-109

Query Match 100.0%; Score 36; DB 4; Length 7;  
Best Local Similarity 100.0%; Pred. No. 1.9e+05;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GQTRSP1 7  
 Db 1 GQTRSP1 7

RESULT 2  
 US-09-177-776-109  
 ; Sequence 109, Application US/09177776A  
 ; Patent No. 6297051

GENERAL INFORMATION:  
 APPLICANT: Zavada, Jan  
 APPLICANT: Pastorekova, Silvia  
 APPLICANT: Pastorek, Jaromir  
 TITLE OF INVENTION: NN Gene and Protein  
 FILE REFERENCE: D-0021.5A

CURRENT APPLICATION NUMBER: US/09/177,776A  
 CURRENT FILING DATE: 1998-10-23  
 EARLIER APPLICATION NUMBER: 08/787,739  
 EARLIER FILING DATE: 1997-01-24  
 EARLIER APPLICATION NUMBER: 08/485,049  
 EARLIER FILING DATE: 1995-06-07  
 EARLIER APPLICATION NUMBER: 08/486,756  
 EARLIER FILING DATE: 1995-06-07  
 EARLIER APPLICATION NUMBER: 08/477,504  
 EARLIER FILING DATE: 1995-06-07  
 EARLIER APPLICATION NUMBER: 08/481,658  
 EARLIER FILING DATE: 1995-06-07  
 EARLIER APPLICATION NUMBER: 08/485,862  
 EARLIER FILING DATE: 1995-06-07  
 EARLIER APPLICATION NUMBER: 08/485,863  
 EARLIER FILING DATE: 1995-06-07  
 EARLIER APPLICATION NUMBER: 08/487,077  
 EARLIER FILING DATE: 1995-06-07  
 EARLIER APPLICATION NUMBER: 08/260,190  
 EARLIER FILING DATE: 1994-06-15  
 EARLIER APPLICATION NUMBER: 08/177,093  
 EARLIER FILING DATE: 1993-12-30  
 EARLIER APPLICATION NUMBER: 07/964,589  
 EARLIER FILING DATE: 1992-10-21  
 EARLIER APPLICATION NUMBER: PV-709-92  
 EARLIER FILING DATE: 1992-03-11  
 NUMBER OF SEQ ID NOS: 116  
 SOFTWARE: PatentIn Ver. 2.0  
 SEQ ID NO 109  
 LENGTH: 7  
 TYPE: PRT  
 ORGANISM: HUMAN  
 US-09-177-776-109

Query Match 100.0%; Score 36; DB 4; Length 7;  
 Best Local Similarity 100.0%; Pred. NO. 1.9e+05;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GQTRSP1 7  
 Db 1 GQTRSP1 7

RESULT 3  
 US-09-254-325-2  
 ; Sequence 2, Application US/09254325  
 ; Patent No. 6090607

GENERAL INFORMATION:  
 APPLICANT:  
 TITLE OF INVENTION: ENHANCED EXPRESSION OF  
 TITLE OF INVENTION: PROTEOLYTIC ENZYMES IN KOJI MOLDS  
 NUMBER OF SEQUENCES: 17  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30 (BPO)  
 CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/254,325  
 FILING DATE:  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 853 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 FEATURE:  
 NAME/KEY: Binding-site  
 LOCATION: 652-676  
 OTHER INFORMATION: /note= "DNA BINDING SITE"  
 NAME/KEY: Region  
 LOCATION: 1..731  
 OTHER INFORMATION: /note= "TRUNCATED AREA WHICH IS  
 STILL ACTIVE BUT NOT REPRESSIONED BY  
 OTHER INFORMATION: L-GLUTAM..."

Query Match 86.1%; Score 31; DB 3; Length 853;  
 Best Local Similarity 71.4%; Pred. NO. 96;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 GQTRSP1 7  
 Db 780 GQTRSP1 786

RESULT 4  
 US-09-079-415-6  
 ; Sequence 6, Application US/09079415  
 ; Patent No. 6013452

GENERAL INFORMATION:  
 APPLICANT: Christensen, Tove  
 APPLICANT: Lembeck, Jan  
 TITLE OF INVENTION: A Fungus Wherein The area, pepc and/or  
 TITLE OF INVENTION: pepc Genes Have Been Inactivated  
 NUMBER OF SEQUENCES: 12  
 CORRESPONDENCE ADDRESSES:  
 ADDRESSEE: No. 6013452a No. 6013452abx of No. 6013452th America, Inc.  
 STREET: 405 Lexington Avenue  
 CITY: New York  
 STATE: NY  
 COUNTRY: USA  
 ZIP: 10174

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette  
 COMPUTER: IBM compatible  
 OPERATING SYSTEM: DOS  
 SOFTWARE: FASTSEQ for windows Version 2.0

CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/079,415  
 FILING DATE: 14-MAY-1998  
 CLASSIFICATION:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Rozek, Carol  
 REGISTRATION NUMBER: 36,993  
 REFERENCE/DOCKET NUMBER: 4657.204-US  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 212-867-0123  
 TELEFAX: 212-878-9655

INFORMATION FOR SEQ ID NO: 6:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 866 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-09-079-415-6

Query Match 86.1%; Score 31; DB 3; Length 866;  
 Best Local Similarity 71.4%; Pred. NO. 97;

Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GQTRSP 7  
 |||||:  
 Db 793 GQTRNPI 799

RESULT 5  
 US-08-750-458A-2  
 ; Sequence 2, Application US/08750458A  
 ; Patent No. 6025185  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Christensen, Tove  
 ; TITLE OF INVENTION: A Fungus Wherein The Area Gene Has Been Modified And An Area  
 ; NUMBER OF SEQUENCES: 2  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: No. 60251850 No. 6025185disk of No. 6025185ch America, Inc.  
 ; STREET: 405 Lexington Avenue, Suite 6400  
 ; CITY: New York  
 ; STATE: New York  
 ; COUNTRY: U.S.A.  
 ; ZIP: 10174-6401  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/750,458A  
 ; FILING DATE: 3-December-1996  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Rozek, Carol E.  
 ; REGISTRATION NUMBER: 36,993  
 ; REFERENCE/DOCKET NUMBER: 4129.204-US  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 212 867 0123  
 ; TELEFAX: 212 867 0298  
 ; INFORMATION FOR SEQ ID NO: 2:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 866 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-08-750-458A-2

Query Match 86.1%; Score 31; DB 3; Length 866;  
 Best Local Similarity 71.4%; Pred. No. 97;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GQTRSP 7  
 |||||:  
 Db 793 GQTRNPI 799

RESULT 6  
 US-09-178-115-107  
 ; Sequence 107, Application US/09178115  
 ; Patent No. 6297041  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Zavada, Jan  
 ; APPLICANT: Pastorekova, Silvia  
 ; APPLICANT: Pastorek, Jaromir  
 ; TITLE OF INVENTION: MN Gene and Protein  
 ; FILE REFERENCE: D-0021.5A  
 ; CURRENT APPLICATION NUMBER: US/09/178,115  
 ; EARLIER FILING DATE: 1998-10-23  
 ; EARLIER APPLICATION NUMBER: 09/177,776  
 ; EARLIER FILING DATE: 1998-10-23  
 ; EARLIER APPLICATION NUMBER: 08/787,739  
 ; EARLIER FILING DATE: 1997-01-24  
 ; EARLIER APPLICATION NUMBER: 08/485,049

EARLIER FILING DATE: 1995-06-07  
 ; EARLIER APPLICATION NUMBER: 08/486,756  
 ; EARLIER FILING DATE: 1995-06-07  
 ; EARLIER APPLICATION NUMBER: 08/477,504  
 ; EARLIER FILING DATE: 1995-06-07  
 ; EARLIER APPLICATION NUMBER: 08/481,658  
 ; EARLIER FILING DATE: 1995-06-07  
 ; EARLIER APPLICATION NUMBER: 08/485,862  
 ; EARLIER FILING DATE: 1995-06-07  
 ; EARLIER APPLICATION NUMBER: 08/485,863  
 ; EARLIER FILING DATE: 1995-06-07  
 ; EARLIER APPLICATION NUMBER: 08/487,077  
 ; EARLIER FILING DATE: 1995-06-07  
 ; EARLIER APPLICATION NUMBER: 08/260,190  
 ; EARLIER FILING DATE: 1994-06-15  
 ; EARLIER APPLICATION NUMBER: 08/177,093  
 ; EARLIER FILING DATE: 1993-12-30  
 ; EARLIER APPLICATION NUMBER: 07/964,589  
 ; EARLIER FILING DATE: 1992-10-21  
 ; EARLIER APPLICATION NUMBER: PV-709-92

Query Match 83.3%; Score 30; DB 4; Length 7;  
 Best Local Similarity 71.4%; Pred. No. 1.9e+05;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GQTRSP 7  
 |||||:  
 Db 1 GQTRAPL 7

RESULT 7  
 US-09-177-776-107  
 ; Sequence 107, Application US/0917776A  
 ; Patent No. 6297051  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Zavada, Jan  
 ; APPLICANT: Pastorekova, Silvia  
 ; APPLICANT: Pastorek, Jaromir  
 ; TITLE OF INVENTION: MN Gene and Protein  
 ; FILE REFERENCE: D-0021.5A  
 ; CURRENT APPLICATION NUMBER: US/09/177,776A  
 ; EARLIER FILING DATE: 1998-10-23  
 ; EARLIER APPLICATION NUMBER: 08/787,739  
 ; EARLIER FILING DATE: 1997-01-24  
 ; EARLIER APPLICATION NUMBER: 08/485,049  
 ; EARLIER FILING DATE: 1995-06-07  
 ; EARLIER APPLICATION NUMBER: 08/486,756  
 ; EARLIER FILING DATE: 1995-06-07  
 ; EARLIER APPLICATION NUMBER: 08/477,504  
 ; EARLIER FILING DATE: 1995-06-07  
 ; EARLIER APPLICATION NUMBER: 08/481,658  
 ; EARLIER FILING DATE: 1995-06-07  
 ; EARLIER APPLICATION NUMBER: 08/485,862  
 ; EARLIER FILING DATE: 1995-06-07  
 ; EARLIER APPLICATION NUMBER: 08/485,863  
 ; EARLIER FILING DATE: 1995-06-07  
 ; EARLIER APPLICATION NUMBER: 08/487,077  
 ; EARLIER FILING DATE: 1995-06-07  
 ; EARLIER APPLICATION NUMBER: 08/260,190  
 ; EARLIER FILING DATE: 1994-06-15  
 ; EARLIER APPLICATION NUMBER: 08/177,093  
 ; EARLIER FILING DATE: 1993-12-30  
 ; EARLIER APPLICATION NUMBER: 07/964,589  
 ; EARLIER FILING DATE: 1992-10-21  
 ; EARLIER APPLICATION NUMBER: PV-709-92

EARLIER FILING DATE: 1992-03-11  
NUMBER OF SEQ ID NOS: 116  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 107  
LENGTH: 7  
TYPE: PRT  
ORGANISM: HUMAN  
US-09-177-776-107

Query Match 83.3%; Score 30; DB 4; Length 7;  
Best Local Similarity 71.4%; Pred. No. 1.9e+05;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 GQTRSPL 7  
Db 1 GETRAPL 7

RESULT 8  
US-09-084-605B-29  
Sequence 29, Application US/09084605B  
Patent No. 6329501  
GENERAL INFORMATION:  
APPLICANT: Smith, Bruce F.  
APPLICANT: Samoilova, Tatiana  
TITLE OF INVENTION: Methods and Compositions for Targeting  
FILE REFERENCE: 5721-8  
CURRENT APPLICATION NUMBER: US/09/084,605B  
CURRENT FILING DATE: 1998-05-26  
NUMBER OF SEQ ID NOS: 29  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 29  
LENGTH: 7  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Phage display library peptides  
US-09-084-605B-29

Query Match 83.3%; Score 30; DB 4; Length 7;  
Best Local Similarity 71.4%; Pred. No. 1.9e+05;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 GQTRSPL 7  
Db 1 GETRAPL 7

RESULT 9  
US-09-438-150-6  
Sequence 6, Application US/09438150  
Patent No. 639575  
GENERAL INFORMATION:  
APPLICANT: Smith, Bruce F.  
APPLICANT: Samoilova, Tatiana I.  
APPLICANT: Baker, Henry J.  
TITLE OF INVENTION: Methods and Compositions for Targeting  
OF INVENTION: Compounds to the Central Nervous System  
FILE REFERENCE: 5721-13  
CURRENT APPLICATION NUMBER: US/09/438,150  
CURRENT FILING DATE: 1999-11-10  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 6  
LENGTH: 7  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Phage display library peptides  
US-09-438-150-6

Query Match 83.3%; Score 30; DB 4; Length 7;

Best Local Similarity 71.4%; Pred. No. 1.9e+05;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 GQTRSPL 7  
Db 1 GETRAPL 7

RESULT 10  
US-09-134-001C-2850  
Sequence 2850, Application US/09134001C  
Patent No. 6380370  
GENERAL INFORMATION:  
APPLICANT: Lynn Doucette-Stamm et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
OF INVENTION: EPIDERMIDS FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: CTC-007  
CURRENT APPLICATION NUMBER: US/09/134,001C  
CURRENT FILING DATE: 1998-08-13  
PRIOR APPLICATION NUMBER: US 60/064,964  
PRIOR FILING DATE: 1997-11-08  
PRIOR APPLICATION NUMBER: US 60/055,779  
PRIOR FILING DATE: 1997-08-14  
NUMBER OF SEQ ID NOS: 5674  
SEQ ID NO 2850  
LENGTH: 295  
TYPE: PRT  
ORGANISM: Staphylococcus epidermidis  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (169)  
OTHER INFORMATION: Identity of amino acid at the above locations are unknown.  
US-09-134-001C-2850

Query Match 83.3%; Score 30; DB 4; Length 295;  
Best Local Similarity 71.4%; Pred. No. 51;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GQTRSPL 7  
Db 105 GQTRGPT 111

RESULT 11  
US-09-178-115-108  
Sequence 108, Application US/09178115  
Patent No. 6297041  
GENERAL INFORMATION:  
APPLICANT: Zavada, Jan  
APPLICANT: Pastorekova, Silvia  
APPLICANT: Pastorek, Jaromir  
TITLE OF INVENTION: MN Gene and Protein  
FILE REFERENCE: D-0021, 5A  
CURRENT APPLICATION NUMBER: US/09/178,115  
CURRENT FILING DATE: 1998-10-23  
EARLIER APPLICATION NUMBER: 09/177,776  
EARLIER FILING DATE: 1998-10-23  
EARLIER APPLICATION NUMBER: 08/787,739  
EARLIER FILING DATE: 1997-01-24  
EARLIER APPLICATION NUMBER: 08/485,049  
EARLIER FILING DATE: 1995-06-07  
EARLIER APPLICATION NUMBER: 08/486,756  
EARLIER FILING DATE: 1995-06-07  
EARLIER APPLICATION NUMBER: 08/477,504  
EARLIER FILING DATE: 1995-06-07  
EARLIER APPLICATION NUMBER: 08/481,658  
EARLIER FILING DATE: 1995-06-07  
EARLIER APPLICATION NUMBER: 08/485,862  
EARLIER FILING DATE: 1995-06-07  
EARLIER APPLICATION NUMBER: 08/485,863  
EARLIER FILING DATE: 1995-06-07  
EARLIER APPLICATION NUMBER: 08/487,077  
EARLIER FILING DATE: 1995-06-07



EARLIER APPLICATION NUMBER: 08/260,190  
EARLIER FILING DATE: 1994-06-15  
EARLIER APPLICATION NUMBER: 08/177,093  
EARLIER FILING DATE: 1993-12-30  
EARLIER APPLICATION NUMBER: 07/964,589  
EARLIER FILING DATE: 1992-10-21  
EARLIER APPLICATION NUMBER: PV-709-92  
EARLIER FILING DATE: 1992-03-11  
NUMBER OF SEQ ID NOS: 116  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO: 108  
LENGTH: 7  
TYPE: PRT  
ORGANISM: HUMAN  
US-09-178-115-108

Query Match 80.6%; Score 29; DB 4; Length 7;  
Best Local Similarity 71.4%; Pred. No. 1.9e+05;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GOTSPL 7  
Db 1 GETREPL 7

RESULT 12  
US-09-177-776-108  
Sequence 108, Application US/09177776A  
Patent No. 6297051  
GENERAL INFORMATION:  
APPLICANT: Zavada, Jan  
APPLICANT: Pastorek, Jaromir  
TITLE OF INVENTION: MN Gene and Protein  
FILE REFERENCE: D-0021.5A  
CURRENT APPLICATION NUMBER: US/09/177,776A  
FILING DATE: 1998-10-23  
EARLIER APPLICATION NUMBER: 08/787,739  
EARLIER FILING DATE: 1997-01-24  
EARLIER APPLICATION NUMBER: 08/485,049  
EARLIER FILING DATE: 1995-06-07  
EARLIER APPLICATION NUMBER: 08/486,756  
EARLIER FILING DATE: 1995-06-07  
EARLIER APPLICATION NUMBER: 08/477,504  
EARLIER FILING DATE: 1995-06-07  
EARLIER APPLICATION NUMBER: 08/481,658  
EARLIER FILING DATE: 1995-06-07  
EARLIER APPLICATION NUMBER: 08/485,862  
EARLIER FILING DATE: 1995-06-07  
EARLIER APPLICATION NUMBER: 08/485,863  
EARLIER FILING DATE: 1995-06-07  
EARLIER APPLICATION NUMBER: 08/487,077  
EARLIER FILING DATE: 1995-06-07  
EARLIER APPLICATION NUMBER: 08/260,190  
EARLIER FILING DATE: 1994-06-15  
EARLIER APPLICATION NUMBER: 08/177,093  
EARLIER FILING DATE: 1993-12-30  
EARLIER APPLICATION NUMBER: 07/964,589  
EARLIER FILING DATE: 1992-10-21  
EARLIER APPLICATION NUMBER: PV-709-92  
EARLIER FILING DATE: 1992-03-11  
NUMBER OF SEQ ID NOS: 116  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO: 108  
LENGTH: 7  
TYPE: PRT  
ORGANISM: HUMAN  
US-09-177-776-108

Query Match 80.6%; Score 29; DB 4; Length 7;  
Best Local Similarity 71.4%; Pred. No. 1.9e+05;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GOTSPL 7  
Db 1 GETREPL 7

RESULT 13  
US-08-044-621D-36  
Sequence 36, Application US/08044621D  
Patent No. 5405769  
GENERAL INFORMATION:  
APPLICANT: Warren W. Makarchuk  
APPLICANT: Wang L. Sung  
APPLICANT: Makoto Yasuchi  
APPLICANT: Robert L. Campbell  
APPLICANT: David R. Rose  
TITLE OF INVENTION: CONSTRUCTION OF THERMOSTABLE MUTANTS  
OF A LOW MOLECULAR MASS XYLANASE  
NUMBER OF SEQUENCES: 37  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Gowling, Strachy & Henderson  
STREET: Suite 2600, 160 Elgin Street  
CITY: Ottawa  
STATE: Ontario  
COUNTRY: Canada  
ZIP: K1P 1C3  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 5.25 in., 360KB storage  
COMPUTER: IBM PC  
OPERATING SYSTEM: PC-DOS  
SOFTWARE: WordPerfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/044,621D  
FILING DATE: April 8, 1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
ATTORNEY/AGENT INFORMATION:  
FILING DATE:  
NAME: Judy A. Erract  
REGISTRATION NUMBER: 34,076  
REFERENCE/DOCKET NUMBER: 08-863796  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 613-786-0199  
TELEFAX: 613-563-9869  
TELEX:  
INFORMATION FOR SEQ ID NO: 36:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 185  
TYPE: Amino Acid  
STRANDEDNESS: No. 5405769 Relevant  
TOPOLOGY: linear  
MOLECULE TYPE:  
DESCRIPTION: protein  
HYPOTHETICAL: No  
ANTI-SENSE: No  
FRAGMENT TYPE: No  
ORIGINAL SOURCE:  
ORGANISM: Bacillus circulans  
IMMEDIATE SOURCE:  
POSITION IN GENOME:  
FEATURE:  
PUBLICATION INFORMATION:  
AUTHORS: Yang R.C.A., Mackenzie C.R. & Narang  
AUTHORS: R.A.  
TITLE:  
JOURNAL: Nucleic Acids Res.  
VOLUME: 16  
ISSUE:  
PAGES: 7187  
DATE: 1988  
DOCUMENT NUMBER:  
FILING DATE:  
PUBLICATION DATE:

RELEVANT RESIDUES IN SEQ ID NO:  
US-08-044-621D-36

Query Match 80.6%; Score 29; DB 1; Length 185;  
Best Local Similarity 85.7%; Pred. No. 51;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GQTRSPL 7  
|  
70 GQTRSPL 76

## RESULT 14

US-08-044-621D-37  
; Sequence 37, Application US/08044621D  
; Patent No. 5405769  
; GENERAL INFORMATION:  
; APPLICANT: Warren W. Wakarchuk  
; APPLICANT: Ming L. Sung  
; APPLICANT: Makoto Yaguchi  
; APPLICANT: Robert L. Campbell  
; APPLICANT: David R. Rose  
; TITLE OF INVENTION: CONSTRUCTION OF THERMOSTABLE MUTANTS  
; TITLE OF INVENTION: OF A LOW MOLECULAR MASS XYLANASE  
; NUMBER OF SEQUENCES: 37  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Gowling, Strathby & Henderson  
; STREET: Suite 2600, 160 Elgin Street  
; CITY: Ottawa  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: K1P 1C3  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 5.25 in., 360KB storage  
; COMPUTER: IBM PC  
; OPERATING SYSTEM: PC-DOS  
; SOFTWARE: WordPerfect 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/044,621D  
; FILING DATE: April 8, 1993  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Judy A. Erratt  
; REGISTRATION NUMBER: 34,076  
; REFERENCE/DOCKET NUMBER: 08-863796  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 613-786-0199  
; TELEFAX: 613-563-9869  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 37:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 185  
; TYPE: Amino Acid  
; STRANDEDNESS: No. 5405769 Relevant  
; TOPOLOGY: linear  
; MOLECULE TYPE:  
; DESCRIPTION: protein  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; FRAGMENT TYPE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Bacillus subtilis  
; IMMEDIATE SOURCE:  
; POSITION IN GENOME:  
; FEATURE:  
; PUBLICATION INFORMATION:  
; AUTHORS: Palce M.G., Bourdonnais R., Desrochers  
; AUTHORS: M., Jurasek L., & Yaguchi M.  
; TITLE:  
; JOURNAL: Arch. Microbiol.

VOLUME: 144

ISSUE:

PAGES: 201-206

DATE: 1986

DOCUMENT NUMBER:

FILING DATE:

PUBLICATION DATE:

RELEVANT RESIDUES IN SEQ ID NO:  
US-08-044-621D-37

Query Match 80.6%; Score 29; DB 1; Length 185;  
Best Local Similarity 85.7%; Pred. No. 51;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GQTRSPL 7  
|  
70 GQTRSPL 76

## RESULT 15

US-08-709-912-3  
; Sequence 3, Application US/08709912  
; Patent No. 5739840  
; GENERAL INFORMATION:  
; APPLICANT: Sung Dr., Wing L  
; APPLICANT: Yaguchi Dr., Makoto  
; APPLICANT: Ishikawa Dr., Kazuhiko  
; TITLE OF INVENTION: Modification of xylanase to improve  
; TITLE OF INVENTION: Thermophilicity, Alkalophilicity and  
; NUMBER OF SEQUENCES: 54  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fitzpatrick, Cella, Harper, and Scinto  
; STREET: 277 Park Ave.  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10172-0194  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/709,912  
; FILING DATE: 09-SEP-1996  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Olsen Mr, Warren E  
; REGISTRATION NUMBER: 27290  
; REFERENCE/DOCKET NUMBER: 1039.2000  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 758-2400  
; TELEFAX: (212) 758-2982  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 185 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; FRAGMENT TYPE: internal  
; ORIGINAL SOURCE:  
; ORGANISM: Bacillus circulans  
; PUBLICATION INFORMATION:  
; AUTHORS: Yang, R.C.A.  
; AUTHORS: Mackenzie, C.R.  
; AUTHORS: Narang, S.A.  
; JOURNAL: Nucleic Acid Research  
; VOLUME: 16  
; PAGES: 7187

DATE: 1988  
US-08-709-912-3

Query Match 80.6%; Score 29; DB 1; Length 185;  
Best Local Similarity 85.7%; Pred. No. 51;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 GQTRSP 7  
Db 70 GQTRSP 76

Search completed: January 29, 2003, 14:04:18  
Job time : 8.61539 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model1

Run on: January 29, 2003, 14:00:01 ; Search time 5.38462 Seconds  
(without alignments)  
26.232 Million cell updates/sec

Title: US-09-807-949a-109  
Perfect score: 36  
Sequence: 1 GQTRSPPL 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 122226 seqs, 20178551 residues  
Total number of hits satisfying chosen parameters: 122226

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

Published Applications AA.\*  
1: /cgn2\_6/ptodata/1/pubppa/US08\_NEW\_PUB.pep.\*  
2: /cgn2\_6/ptodata/1/pubppa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubppa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubppa/US07\_NEW\_PUB.pep.\*  
5: /cgn2\_6/ptodata/1/pubppa/US07\_PUBCOMB.pep.\*  
6: /cgn2\_6/ptodata/1/pubppa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubppa/US08\_PUBCOMB.pep.\*  
8: /cgn2\_6/ptodata/1/pubppa/US09\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubppa/US09\_NEW\_PUB.pep.\*  
10: /cgn2\_6/ptodata/1/pubppa/US09\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubppa/US10\_NEW\_PUB.pep.\*  
12: /cgn2\_6/ptodata/1/pubppa/US10\_PUBCOMB.pep.\*  
13: /cgn2\_6/ptodata/1/pubppa/US60\_NEW\_PUB.pep.\*  
14: /cgn2\_6/ptodata/1/pubppa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	88.9	48	10	US-09-864-761-36972	Sequence 36972, A
2	86.1	635	10	US-09-815-242-11209	Sequence 11209, A
3	86.1	637	10	US-09-815-242-12058	Sequence 12058, A
4	86.1	665	10	US-09-942-447-2	Sequence 2, Appl1
5	83.3	7	10	US-09-947-137-29	Sequence 29, Appl1
6	83.3	354	9	US-09-430-029-7	Sequence 7, Appl1
7	83.3	369	10	US-09-815-242-10807	Sequence 10807, A
8	80.6	101	9	US-09-786-692-1971	Sequence 1971, Ap
9	80.6	101	9	US-09-796-692-2487	Sequence 2487, Ap
10	80.6	208	10	US-09-864-761-35865	Sequence 35865, A
11	80.6	416	10	US-09-815-242-11301	Sequence 11301, A
12	80.6	416	10	US-09-881-752A-172	Sequence 172, Appl
13	80.6	637	10	US-09-759-010-1	Sequence 1, Appl1
14	80.6	638	10	US-09-815-242-10015	Sequence 10015, A
15	80.6	638	10	US-09-815-242-13713	Sequence 13713, A
16	80.6	1230	10	US-09-727-384-8	Sequence 8, Appl1
17	80.6	60	9	US-09-738-626-4050	Sequence 4050, Ap
18	77.8	83	10	US-09-764-877-1796	Sequence 1796, Ap
19	77.8	110	10	US-09-925-297-656	Sequence 656, App

20	28	77.8	123	10	US-09-867-550-84	Sequence 84, Appl
21	28	77.8	136	10	US-09-925-299-1246	Sequence 1246, Ap
22	28	77.8	150	9	US-09-764-868-938	Sequence 938, App
23	28	77.8	198	9	US-09-981-353-138	Sequence 138, App
24	28	77.8	231	9	US-09-252-150-16	Sequence 16, Appl
25	28	77.8	231	9	US-09-252-150-17	Sequence 17, Appl
26	28	77.8	234	9	US-09-252-150-21	Sequence 21, Appl
27	28	77.8	246	9	US-09-252-150-18	Sequence 18, Appl
28	28	77.8	258	9	US-09-252-150-19	Sequence 19, Appl
29	28	77.8	250	9	US-09-252-150-20	Sequence 20, Appl
30	28	77.8	322	10	US-09-942-447-5	Sequence 5, Appl1
31	28	77.8	427	10	US-09-942-447-4	Sequence 4, Appl1
32	27	75.0	76	10	US-09-864-761-35027	Sequence 35027, A
33	27	75.0	247	9	US-09-854-133-112	Sequence 112, App
34	27	75.0	247	10	US-09-738-973-112	Sequence 112, App
35	27	75.0	271	10	US-09-925-297-545	Sequence 545, App
36	27	75.0	306	10	US-09-835-788A-19	Sequence 19, Appl
37	27	75.0	364	9	US-09-764-868-1087	Sequence 1087, Ap
38	27	75.0	386	9	US-09-764-868-1011	Sequence 1011, Ap
39	27	75.0	438	10	US-09-815-242-11509	Sequence 11509, A
40	27	75.0	446	10	US-09-815-242-12338	Sequence 12338, A
41	27	75.0	446	10	US-09-815-242-12775	Sequence 12775, A
42	27	75.0	448	10	US-09-815-242-13378	Sequence 13378, A
43	27	75.0	451	10	US-09-815-242-10514	Sequence 10514, A
44	27	75.0	451	10	US-09-815-242-10514	Sequence 10514, A
45	27	75.0	612	9	US-09-764-868-663	Sequence 663, App

#### ALIGNMENTS

RESULT 1  
US-09-864-761-36972  
; Sequence 36972, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharon G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; FILE REFERENCE: Aeonica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30

```
/ PRIOR APPLICATION NUMBER: PCT/US01/00670
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: US 60/234,687
/ PRIOR FILING DATE: 2000-09-21
/ PRIOR APPLICATION NUMBER: US 09/608,408
/ PRIOR FILING DATE: 2000-06-30
/ PRIOR APPLICATION NUMBER: US 09/774,203
/ PRIOR FILING DATE: 2001-01-29
/ NUMBER OF SEQ ID NOS: 49117
/ SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
/ SEQ ID NO 36972
/ LENGTH: 48
/ TYPE: PRT
/ ORGANISM: Homo sapiens
/ FEATURE:
/ OTHER INFORMATION: MAP TO AL023804.1
/ OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 0.79
/ OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.1
/ OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 0.98
/ OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.93
/ OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.68
/ OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.92
/ OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.98
/ OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.2
/ OTHER INFORMATION: EXPRESSED IN HB1100, SIGNAL = 2.2
/ OTHER INFORMATION: EST_HUMAN HIT: A1688948.1, EVALU8 4.50e-01
/ US-09-864-761-36972
```

```
Query Match      88.9%; Score 32; DB 10; Length 48;
Best Local Similarity 100.0%; Pred. No. 2.3;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 GQTRSP 6
        |||||
DB      10 GQTRSP 15
```

```
RESULT 2
US-09-815-242-11209
/ Sequence 11209, Application US/09815242
/ Patent No. US20020061569A1
/ GENERAL INFORMATION:
/ APPLICANT: Haselbeck, Robert
/ APPLICANT: Ohlsen, Karl L.
/ APPLICANT: Zykkind, Judith W.
/ APPLICANT: Wall, Daniel
/ APPLICANT: Trawick, John D.
/ APPLICANT: Carr, Grant J.
/ APPLICANT: Yamamoto, Robert T.
/ APPLICANT: Xu, H. Howard
/ TITLE OF INVENTION: Identification of Essential Genes in
/ TITLE OF INVENTION: Prokaryotes
/ FILE REFERENCE: ELITRA.011A
/ CURRENT APPLICATION NUMBER: US/09/815,242
/ CURRENT FILING DATE: 2001-03-21
/ PRIOR APPLICATION NUMBER: 60/191,078
/ PRIOR FILING DATE: 2000-03-21
/ PRIOR APPLICATION NUMBER: 60/206,848
/ PRIOR FILING DATE: 2000-05-23
/ PRIOR APPLICATION NUMBER: 60/207,727
/ PRIOR FILING DATE: 2000-05-26
/ PRIOR APPLICATION NUMBER: 60/242,578
/ PRIOR FILING DATE: 2000-10-23
/ PRIOR APPLICATION NUMBER: 60/253,625
/ PRIOR FILING DATE: 2000-11-27
/ PRIOR APPLICATION NUMBER: 60/257,931
/ PRIOR FILING DATE: 2000-12-22
/ PRIOR APPLICATION NUMBER: 60/269,308
/ PRIOR FILING DATE: 2001-02-16
/ NUMBER OF SEQ ID NOS: 14110
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 11209
/ LENGTH: 635
```

```
/ TYPE: PRT
/ ORGANISM: Haemophilus influenzae
/ US-09-815-242-11209
```

```
Query Match      86.1%; Score 31; DB 10; Length 635;
Best Local Similarity 85.7%; Pred. No. 59;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY      1 GQTRSP 7
        |||||
DB      341 GQTRMPL 347
```

```
RESULT 3
US-09-815-242-12058
/ Sequence 12058, Application US/09815242
/ Patent No. US20020061569A1
/ GENERAL INFORMATION:
/ APPLICANT: Haselbeck, Robert
/ APPLICANT: Ohlsen, Karl L.
/ APPLICANT: Zykkind, Judith W.
/ APPLICANT: Wall, Daniel
/ APPLICANT: Trawick, John D.
/ APPLICANT: Carr, Grant J.
/ APPLICANT: Yamamoto, Robert T.
/ APPLICANT: Xu, H. Howard
/ TITLE OF INVENTION: Identification of Essential Genes in
/ TITLE OF INVENTION: Prokaryotes
/ FILE REFERENCE: ELITRA.011A
/ CURRENT APPLICATION NUMBER: US/09/815,242
/ CURRENT FILING DATE: 2001-03-21
/ PRIOR APPLICATION NUMBER: 60/191,078
/ PRIOR FILING DATE: 2000-03-21
/ PRIOR APPLICATION NUMBER: 60/206,848
/ PRIOR FILING DATE: 2000-05-23
/ PRIOR APPLICATION NUMBER: 60/207,727
/ PRIOR FILING DATE: 2000-05-26
/ PRIOR APPLICATION NUMBER: 60/242,578
/ PRIOR FILING DATE: 2000-10-23
/ PRIOR APPLICATION NUMBER: 60/253,625
/ PRIOR FILING DATE: 2000-11-27
/ PRIOR APPLICATION NUMBER: 60/257,931
/ PRIOR FILING DATE: 2000-12-22
/ PRIOR APPLICATION NUMBER: 60/269,308
/ PRIOR FILING DATE: 2001-02-16
/ NUMBER OF SEQ ID NOS: 14110
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 12058
/ LENGTH: 637
/ TYPE: PRT
/ ORGANISM: Pseudomonas aeruginosa
/ US-09-815-242-12058
```

```
US-09-815-242-12058
```

```
Query Match      86.1%; Score 31; DB 10; Length 637;
Best Local Similarity 85.7%; Pred. No. 59;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY      1 GQTRSP 7
        |||||
DB      342 GQTRMPL 348
```

```
RESULT 4
US-09-942-447-2
/ Sequence 2, Application US/09942447
/ Patent No. US20020127567A1
/ GENERAL INFORMATION:
/ APPLICANT: Millennium Pharmaceuticals, Inc.
/ APPLICANT: Glucksmann, Maria A.
/ TITLE OF INVENTION: 52991, A NOVEL HUMAN TRANSPORTER AND USES
/ THEREFOR
/ FILE REFERENCE: 38155-20032.00
/ CURRENT APPLICATION NUMBER: US/09/942,447
```

;; CURRENT FILING DATE: 2002-01-16  
;; PRIOR APPLICATION NUMBER: US 60/229,036  
;; PRIOR FILING DATE: 2000-08-31  
;; NUMBER OF SEQ ID NOS: 9  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 2  
;; LENGTH: 665  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
US-09-942-447-2

Query Match 86.1%; Score 31; DB 10; Length 665;  
Best Local Similarity 85.7%; Pred. No. 62;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GQTRSP 7  
DB 314 GQTRSP 320

RESULT 5  
US-09-947-137-29  
; Sequence 29, Application US/09947137  
; Patent No. US20020137023A1  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Bruce F.  
; APPLICANT: Samoilova, Tatiana  
; TITLE OF INVENTION: Methods and Compositions for Targeting  
; FILE REFERENCE: 5721-8  
; CURRENT APPLICATION NUMBER: US/09/947,137  
; CURRENT FILING DATE: 2001-09-05  
; PRIOR APPLICATION NUMBER: 09/084,605  
; PRIOR FILING DATE: 1998-05-26  
; NUMBER OF SEQ ID NOS: 29  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 29  
; LENGTH: 7  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURES:  
; OTHER INFORMATION: Phage display library peptides  
US-09-947-137-29

Query Match 83.3%; Score 30; DB 10; Length 7;  
Best Local Similarity 71.4%; Pred. No. 1e+05;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 GQTRSP 7  
DB 1 GQTRAP 7

RESULT 6  
US-09-430-029-7  
; Sequence 7, Application US/09430029  
; Patent No. US20020168738A1  
; GENERAL INFORMATION:  
; APPLICANT: Yano, Tetsuya; No. US20020168738A1oto, tsuyoshi; Imamura, Takeshi; Canon K  
; TITLE OF INVENTION: DNA Fragment Carrying Toluene Monooxygenase Gene,  
; TITLE OF INVENTION: Recombinant Plasmid, Transformed Microorganism,  
; TITLE OF INVENTION: Method for Degrading Chlorinated Aliphatic Hydrocarbon  
; TITLE OF INVENTION: Compounds and Aromatic Compounds, and  
; TITLE OF INVENTION: Method for Environmental Remediation  
; FILE REFERENCE: C0013982US  
; CURRENT APPLICATION NUMBER: US/09/430,029  
; CURRENT FILING DATE: 1999-10-29  
; EARLIER APPLICATION NUMBER: JP P1998-310801  
; EARLIER FILING DATE: 1998-10-30  
; NUMBER OF SEQ ID NOS: 13  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 7  
; LENGTH: 354

;; TYPE: PRT  
;; ORGANISM: Burkholderia cepacia  
;; FEATURE:  
;; OTHER INFORMATION: TOMP polypeptide  
US-09-430-029-7

Query Match 83.3%; Score 30; DB 9; Length 354;  
Best Local Similarity 100.0%; Pred. No. 51;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GQTRSP 7  
DB 345 GQTRSP 350

RESULT 7  
US-09-815-242-10807  
; Sequence 10807, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Karl L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; FILE REFERENCE: ELITRA, 011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 10807  
; LENGTH: 369  
; TYPE: PRT  
; ORGANISM: Enterococcus faecalis  
US-09-815-242-10807

Query Match 83.3%; Score 30; DB 10; Length 369;  
Best Local Similarity 71.4%; Pred. No. 53;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GQTRSP 7  
DB 179 GQTRSP 185

RESULT 8  
US-09-796-692-1971  
; Sequence 1971, Application US/09796692  
; Publication No. US20020198362A1  
; GENERAL INFORMATION:  
; APPLICANT: Gaiger, Alexander  
; APPLICANT: Algate, Paul A.  
; APPLICANT: Mannion, Jane  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY

;; TITLE OF INVENTION: HEMATOLOGICAL MALIGNANCIES  
;; FILE REFERENCE: 2077.001200  
;; CURRENT APPLICATION NUMBER: US/09/796,692  
;; CURRENT FILING DATE: 2001-03-01  
;; PRIOR APPLICATION NUMBER: 60/186,126  
;; PRIOR FILING DATE: 2000-03-01  
;; PRIOR APPLICATION NUMBER: 60/190,479  
;; PRIOR FILING DATE: 2000-03-17  
;; PRIOR APPLICATION NUMBER: 60/200,545  
;; PRIOR FILING DATE: 2000-04-27  
;; PRIOR APPLICATION NUMBER: 60/200,303  
;; PRIOR FILING DATE: 2000-04-28  
;; PRIOR APPLICATION NUMBER: 60/200,779  
;; PRIOR FILING DATE: 2000-04-28  
;; PRIOR APPLICATION NUMBER: 60/200,999  
;; PRIOR FILING DATE: 2000-05-01  
;; PRIOR APPLICATION NUMBER: 60/202,084  
;; PRIOR FILING DATE: 2000-05-04  
;; PRIOR APPLICATION NUMBER: 60/206,201  
;; PRIOR FILING DATE: 2000-05-22  
;; PRIOR APPLICATION NUMBER: 60/218,950  
;; PRIOR FILING DATE: 2000-07-14  
;; PRIOR APPLICATION NUMBER: 60/222,903  
;; PRIOR FILING DATE: 2000-08-03  
;; PRIOR APPLICATION NUMBER: 60/223,416  
;; PRIOR FILING DATE: 2000-08-04  
;; PRIOR APPLICATION NUMBER: 60/223,378  
;; PRIOR FILING DATE: 2000-08-07  
;; NUMBER OF SEQ ID NOS: 9597  
;; SOFTWARE: FastSeq for Windows Version 3.0  
;; SEQ ID NO 1971  
;; LENGTH: 101  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
US-09-796-692-1971

Query Match 80.6%; Score 29; DB 9; Length 101;  
Best Local Similarity 83.3%; Pred. No. 22;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GQTRSP 6  
Db 74 GQTRAP 79

RESULT 9  
US-09-796-692-2487  
;; Sequence 2487, Application US/09796692  
;; Publication No. US20020198362A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Gaigier, Alexander  
;; APPLICANT: Aigate, Paul A.  
;; APPLICANT: Mannion, Jane  
;; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY  
;; FILE REFERENCE: 2077.001200  
;; CURRENT APPLICATION NUMBER: US/09/796,692  
;; CURRENT FILING DATE: 2001-03-01  
;; PRIOR APPLICATION NUMBER: 60/186,126  
;; PRIOR FILING DATE: 2000-03-01  
;; PRIOR APPLICATION NUMBER: 60/190,479  
;; PRIOR FILING DATE: 2000-03-17  
;; PRIOR APPLICATION NUMBER: 60/200,545  
;; PRIOR FILING DATE: 2000-04-27  
;; PRIOR APPLICATION NUMBER: 60/200,303  
;; PRIOR FILING DATE: 2000-04-28  
;; PRIOR APPLICATION NUMBER: 60/200,779  
;; PRIOR FILING DATE: 2000-04-28  
;; PRIOR APPLICATION NUMBER: 60/200,999  
;; PRIOR FILING DATE: 2000-05-01  
;; PRIOR APPLICATION NUMBER: 60/202,084  
;; PRIOR FILING DATE: 2000-05-04  
;; PRIOR APPLICATION NUMBER: 60/206,201

;; PRIOR FILING DATE: 2000-05-22  
;; PRIOR APPLICATION NUMBER: 60/218,950  
;; PRIOR FILING DATE: 2000-07-14  
;; PRIOR APPLICATION NUMBER: 60/222,903  
;; PRIOR FILING DATE: 2000-08-03  
;; PRIOR APPLICATION NUMBER: 60/223,416  
;; PRIOR FILING DATE: 2000-08-04  
;; PRIOR APPLICATION NUMBER: 60/223,378  
;; PRIOR FILING DATE: 2000-08-07  
;; NUMBER OF SEQ ID NOS: 9597  
;; SOFTWARE: FastSeq for Windows Version 3.0  
;; SEQ ID NO 2487  
;; LENGTH: 101  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
US-09-796-692-2487

Query Match 80.6%; Score 29; DB 9; Length 101;  
Best Local Similarity 83.3%; Pred. No. 22;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GQTRSP 6  
Db 74 GQTRAP 79

RESULT 10  
US-09-864-761-35865  
;; Sequence 35865, Application US/09864761  
;; Patent No. US2002048763A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Penn, Sharon G.  
;; APPLICANT: Rank, David R.  
;; APPLICANT: Hanzel, David K.  
;; APPLICANT: Chen, Wenhseng  
;; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
;; FILE REFERENCE: Aeomica-X-1  
;; CURRENT APPLICATION NUMBER: US/09/864,761  
;; CURRENT FILING DATE: 2001-05-23  
;; PRIOR APPLICATION NUMBER: US 60/180,312  
;; PRIOR FILING DATE: 2000-02-04  
;; PRIOR APPLICATION NUMBER: US 60/207,456  
;; PRIOR FILING DATE: 2000-05-26  
;; PRIOR APPLICATION NUMBER: US 09/632,366  
;; PRIOR FILING DATE: 2000-08-03  
;; PRIOR APPLICATION NUMBER: GB 24263.6  
;; PRIOR FILING DATE: 2000-10-04  
;; PRIOR APPLICATION NUMBER: US 60/236,359  
;; PRIOR FILING DATE: 2000-09-27  
;; PRIOR APPLICATION NUMBER: PCT/US01/00666  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00667  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00664  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00669  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00665  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00668  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00663  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00662  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00661  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00670  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: US 60/234,687  
;; PRIOR FILING DATE: 2000-09-21  
;; PRIOR APPLICATION NUMBER: US 09/608,408



PRIOR FILING DATE: 2000-06-30  
 PRIOR APPLICATION NUMBER: US 09/774,203  
 PRIOR FILING DATE: 2001-01-29  
 NUMBER OF SEQ ID NOS: 49117  
 SOFTWARE: Attnomax Sequence Listing Engine vers. 1.1  
 SEQ ID NO 35865  
 LENGTH: 208  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 FEATURE:  
 OTHER INFORMATION: MAP TO AP00504.1  
 OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.3  
 OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.99  
 OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1  
 OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.4  
 OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.1  
 OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.3  
 OTHER INFORMATION: EXPRESSED IN HEL100, SIGNAL = 1.4  
 OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.2  
 OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1  
 OTHER INFORMATION: EST HUMAN HIT: BB795445.1, EVALUATE 9.00e-64  
 OTHER INFORMATION: SWISSPROT HIT: Q13625, EVALUATE 1.70e-02  
 US-09-864-761-35865

Query Match 80.6%; Score 29; DB 10; Length 208;  
 Best Local Similarity 83.3%; Pred. No. 47;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GQTRSP 6  
 Db 201 GETRSP 206

RESULT 11  
 US-09-815-242-11301  
 Sequence 11301, Application US/09815242  
 Patent No. US20020061569A1  
 GENERAL INFORMATION:  
 APPLICANT: Haselbeck, Robert  
 APPLICANT: Ohlsen, Karl L.  
 APPLICANT: Zyskind, Judith W.  
 APPLICANT: Wall, Daniel  
 APPLICANT: Trawick, John D.  
 APPLICANT: Carr, Grant J.  
 APPLICANT: Yamamoto, Robert T.  
 APPLICANT: Xu, H. Howard  
 TITLE OF INVENTION: Identification of Essential Genes in  
 FILE REFERENCE: ELITRA.011A  
 CURRENT APPLICATION NUMBER: US/09/815,242  
 CURRENT FILING DATE: 2001-03-21  
 PRIOR APPLICATION NUMBER: 60/191,078  
 PRIOR FILING DATE: 2000-03-21  
 PRIOR APPLICATION NUMBER: 60/206,848  
 PRIOR FILING DATE: 2000-05-23  
 PRIOR APPLICATION NUMBER: 60/207,727  
 PRIOR FILING DATE: 2000-05-26  
 PRIOR APPLICATION NUMBER: 60/242,578  
 PRIOR FILING DATE: 2000-10-23  
 PRIOR APPLICATION NUMBER: 60/253,625  
 PRIOR FILING DATE: 2000-11-27  
 PRIOR APPLICATION NUMBER: 60/257,931  
 PRIOR FILING DATE: 2000-12-22  
 PRIOR APPLICATION NUMBER: 60/269,308  
 NUMBER OF SEQ ID NOS: 14110  
 SOFTWARE: FastSeq for Windows Version 4.0  
 SEQ ID NO 11301  
 LENGTH: 416  
 TYPE: PRT  
 ORGANISM: Helicobacter pylori  
 US-09-815-242-11301

Query Match 80.6%; Score 29; DB 10; Length 416;  
 Best Local Similarity 83.3%; Pred. No. 98;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GQTRSP 6  
 Db 346 GETRSP 351

RESULT 12  
 US-09-881-752A-172  
 Sequence 172, Application US/09881752A  
 Patent No. US20020115078A1  
 GENERAL INFORMATION:  
 APPLICANT: Kleinhous, Harold  
 APPLICANT: Al-Garawi, Amal  
 APPLICANT: Miller, Charles  
 APPLICANT: Tomb, Jean-Francois  
 APPLICANT: Oomen, Raymond P.  
 TITLE OF INVENTION: Identification of Polynucleotides  
 FILE REFERENCE: Encoding No. US20020115078A1 Helicobacter Polypeptides in the  
 TITLE OF INVENTION: Genome  
 FILE REFERENCE: 06132/041002  
 CURRENT APPLICATION NUMBER: US/09/881,752A  
 CURRENT FILING DATE: 2001-06-15  
 PRIOR APPLICATION NUMBER: US 08/833,457  
 PRIOR FILING DATE: 1997-04-01  
 NUMBER OF SEQ ID NOS: 370  
 SOFTWARE: FastSeq for Windows Version 4.0  
 SEQ ID NO 172  
 LENGTH: 416  
 TYPE: PRT  
 ORGANISM: Helicobacter pylori  
 US-09-881-752A-172

Query Match 80.6%; Score 29; DB 10; Length 416;  
 Best Local Similarity 83.3%; Pred. No. 98;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GQTRSP 6  
 Db 346 GETRSP 351

RESULT 13  
 US-09-759-010-1  
 Sequence 1, Application US/09759010  
 Patent No. US20010034042A1  
 GENERAL INFORMATION:  
 APPLICANT: Srivastava, Pramod K.  
 TITLE OF INVENTION: COMPLEXES OF PEPTIDE BINDING FRAGMENTS OF HEAT-SHOCK  
 FILE REFERENCE: 8449-135  
 CURRENT APPLICATION NUMBER: US/09/759,010  
 CURRENT FILING DATE: 2001-01-12  
 NUMBER OF SEQ ID NOS: 11  
 SOFTWARE: Patentm Ver. 2.1  
 SEQ ID NO 1  
 LENGTH: 637  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 US-09-759-010-1

Query Match 80.6%; Score 29; DB 10; Length 637;  
 Best Local Similarity 71.4%; Pred. No. 1.5e+02;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GQTRSP 7  
 Db 341 GQTRSP 347

RESULT 14

```
US-09-815-242-10015
; Sequence 10015, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; NUMBER OF SEQ ID NOS: 1410
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10015
; LENGTH: 638
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-815-242-10015

Query Match      80.6%; Score 29; DB 10; Length 638;
Best Local Similarity 71.4%; Pred. No. 1.5e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GQTRSP 7
DB 342 GQTRMP 348

RESULT 15
US-09-815-242-13713
; Sequence 13713, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; NUMBER OF SEQ ID NOS: 1410
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13713
; LENGTH: 638
; TYPE: PRT
; ORGANISM: Salmonella typhi
US-09-815-242-13713

Query Match      80.6%; Score 29; DB 10; Length 638;
Best Local Similarity 71.4%; Pred. No. 1.5e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
```

```
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 1410
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13713
; LENGTH: 638
; TYPE: PRT
; ORGANISM: Salmonella typhi
US-09-815-242-13713

Query Match      80.6%; Score 29; DB 10; Length 638;
Best Local Similarity 71.4%; Pred. No. 1.5e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GQTRSP 7
DB 342 GQTRMP 348

Search completed: January 29, 2003, 14:04:55
Job time : 6.38462 secs
```

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2003, 13:51:05 ; Search time 32.5385 Seconds  
(without alignments)  
36.857 Million cell updates/sec

Title: US-09-807-949a-137  
Perfect score: 43  
Sequence: 1 AKMKRRKA 9

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A.Geneseq.101002.\*

1:	/SID2/gcgdata/geneseq/genesep-emb1/AA1980.DAT.*
2:	/SID2/gcgdata/geneseq/genesep-emb1/AA1981.DAT.*
3:	/SID2/gcgdata/geneseq/genesep-emb1/AA1982.DAT.*
4:	/SID2/gcgdata/geneseq/genesep-emb1/AA1983.DAT.*
5:	/SID2/gcgdata/geneseq/genesep-emb1/AA1984.DAT.*
6:	/SID2/gcgdata/geneseq/genesep-emb1/AA1985.DAT.*
7:	/SID2/gcgdata/geneseq/genesep-emb1/AA1986.DAT.*
8:	/SID2/gcgdata/geneseq/genesep-emb1/AA1987.DAT.*
9:	/SID2/gcgdata/geneseq/genesep-emb1/AA1988.DAT.*
10:	/SID2/gcgdata/geneseq/genesep-emb1/AA1989.DAT.*
11:	/SID2/gcgdata/geneseq/genesep-emb1/AA1990.DAT.*
12:	/SID2/gcgdata/geneseq/genesep-emb1/AA1991.DAT.*
13:	/SID2/gcgdata/geneseq/genesep-emb1/AA1992.DAT.*
14:	/SID2/gcgdata/geneseq/genesep-emb1/AA1993.DAT.*
15:	/SID2/gcgdata/geneseq/genesep-emb1/AA1994.DAT.*
16:	/SID2/gcgdata/geneseq/genesep-emb1/AA1995.DAT.*
17:	/SID2/gcgdata/geneseq/genesep-emb1/AA1996.DAT.*
18:	/SID2/gcgdata/geneseq/genesep-emb1/AA1997.DAT.*
19:	/SID2/gcgdata/geneseq/genesep-emb1/AA1998.DAT.*
20:	/SID2/gcgdata/geneseq/genesep-emb1/AA1999.DAT.*
21:	/SID2/gcgdata/geneseq/genesep-emb1/AA2000.DAT.*
22:	/SID2/gcgdata/geneseq/genesep-emb1/AA2001.DAT.*
23:	/SID2/gcgdata/geneseq/genesep-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	43	100.0	9	AA03055	MN protein CA doma
2	35	81.4	69	AB049239	Listeria monocytog
3	34	79.1	482	ABP26198	Streptococcus poly
4	34	79.1	567	ABP57857	Drosophila melanog
5	33	76.7	119	AA009016	Human polypeptide
6	33	76.7	120	AA000681	Human secreted pro
7	33	76.7	120	AA000682	Human secreted pro
8	33	76.7	193	ABG23003	Novel human diagno
9	33	76.7	218	AAU59552	Human G protein-co
10	33	76.7	225	ABP51251	Human MDT SEQ ID

11	33	76.7	505	20	AAV49903
12	33	76.7	569	21	AA048234
13	33	76.7	691	21	AA048233
14	33	76.7	705	21	AA048232
15	32	74.4	221	22	AB065059
16	32	74.4	539	22	AAU32146
17	32	74.4	921	18	AAW22480
18	32	74.4	921	21	AAV77902
19	32	74.4	1049	22	AB058717
20	32	74.4	6815	22	AB066811
21	31	72.1	76	21	AA02062
22	31	72.1	124	22	AA012809
23	31	72.1	135	22	AA001688
24	31	72.1	189	32	AAU87107
25	31	72.1	193	19	AAW59879
26	31	72.1	240	21	AAV52199
27	31	72.1	273	23	AB048624
28	31	72.1	280	20	AAV29606
29	31	72.1	280	22	AA066675
30	31	72.1	398	21	AA029822
31	31	72.1	432	21	AA048078
32	31	72.1	435	21	AA029821
33	31	72.1	469	21	AA048077
34	31	72.1	527	21	AA029820
35	31	72.1	561	21	AA048076
36	31	72.1	580	22	AAW39548
37	31	72.1	592	20	AAV05369
38	31	72.1	592	22	AAW39547
39	31	72.1	592	23	AAU85547
40	31	72.1	648	22	AB063595
41	31	72.1	809	23	AB091315
42	30	69.8	13	21	AAV69736
43	30	69.8	13	22	AAU06073
44	30	69.8	29	16	AA084173
45	30	69.8	35	23	AB066837

## ALIGNMENTS

RESULT 1					
AA03055 standard; peptide; 9 AA.					
XX	XX	AA03055;			
AC	XX	25-SEP-2000 (first entry)			
DT	XX				
XX	XX	MN protein CA domain-binding peptide, SEQ ID NO:137.			
DE	XX				
KW	XX	MN protein; tumour associated cell adhesion molecule; oncoprotein;			
KW	XX	proteoglycan domain; PG domain; carbonic anhydrase; CA domain;			
KW	XX	abnormal expression; neoplastic disease; cancer; gene therapy;			
KW	XX	phage display library.			
OS	XX	Synthetic.			
XX	XX				
PN	XX	W0200024913-A2.			
XX	XX				
PD	XX	04-MAY-2000.			
XX	XX				
PF	XX	22-OCT-1999; 99WC-US24879.			
XX	XX				
PR	XX	23-OCT-1998; 98US-017776.			
XX	XX	23-OCT-1998; 98US-0178115.			
PA	XX	(FARB ) BAYER CORP.			
XX	XX	(VITRO-) INST VIROLOGY.			
PI	XX	Zavada J, Pastorekova S, Pastorek J;			
XX	XX	WPI, 2000-350752/30.			

Mouse ROSA26 antis  
Arabidopsis thalia  
Arabidopsis thalia  
Arabidopsis thalia  
Arabidopsis thalia  
Drosophila melanog  
Novel human secret  
Plasmodium falciparum  
P. falciparum ebl-  
Drosophila melanog  
Drosophila melanog  
Human secreted pro  
Human polypeptide  
Human polypeptide  
Novel central nerv  
Amo acid sequenc  
Human tumour suppr  
Listeria monocytog  
Human INGL protei  
Tumour suppressor  
Arabidopsis thalia  
Arabidopsis thalia  
Arabidopsis thalia  
Arabidopsis thalia  
Arabidopsis thalia  
Human polypeptide  
Human polypeptide  
Clone #20661 (L974  
Drosophila melanog  
Herbicidally activ  
RNA binding peptid  
RNA binding peptid  
Peptide enhancer O  
Human prostate spe

PT A molecule which specifically binds to a site on MN protein  
 PT (oncoprotein) and prevents adhesion of vertebrate cells to the protein.  
 PT useful for treating preneoplastic or neoplastic diseases such as cancer

XX Claim 5, Page 69, 154pp; English.

CC The invention relates to the inhibition of cell adhesion mediated by  
 CC the MN oncoprotein (also known as the MN/CA IX isoenzyme or the MN/G250  
 CC protein). The MN protein is a tumour-associated adhesion molecule which  
 CC comprises a proteoglycan-like (PG) domain (AAB03017) which contains the  
 CC protein's binding site, and a carbonic anhydrase (CA) domain (AAB03018).  
 CC Abnormal expression of the MN protein is associated with tumorigenicity.  
 CC The invention encompasses molecules (e.g., proteins and peptides) which  
 CC which specifically bind to a site on the MN protein, thereby preventing  
 CC adhesion of vertebrate cells to the protein in a cell adhesion assay. It  
 CC also encompasses MN proteins or MN protein fragments which can be added  
 CC to the extracellular environment to prevent the adhesion of vertebrate  
 CC cells to each other. The invention also relates to the identification of  
 CC the binding site of the MN protein and to a method of identifying a site  
 CC on an MN protein to which cells adhere, comprising testing a series of  
 CC overlapping peptides from the protein in a cell adhesion assay. The  
 CC invention encompasses a vector comprising an expression control sequence  
 CC operatively linked to a nucleic acid encoding the variable domains of a  
 CC MN-specific antibody, where the domains are separated by a flexible  
 CC linker peptide (AAB03035) and the vector inhibits the growth of a  
 CC vertebrate preneoplastic or neoplastic cell that abnormally expresses MN  
 CC protein. The invention also encompasses a vector comprising a  
 CC nucleic acid encoding a cytotoxic protein or peptide operatively linked  
 CC to the MN gene promoter, which inhibits the growth of a vertebrate  
 CC preneoplastic or neoplastic cell. Also claimed is a repressor complex,  
 CC that binds to the MN gene promoter (AA52473). MN proteins and peptides,  
 CC MN-binding proteins and peptides, and expression vectors encoding such  
 CC proteins and peptides are useful for treating patients with  
 CC preneoplastic or neoplastic disease (e.g., cancers) associated with or  
 CC characterised by abnormal MN expression. Sequences AAB03032-B03034 and  
 CC AAB03055-B03058 represent synthetic phage display library peptides which  
 CC bind to the CA domain of the human MN protein (AAB03005).

XX Sequence 9 AA;

Query Match 100.0%; Score 43; DB 21; Length 9;  
 Best Local Similarity 100.0%; Pred. No. 7.8e+05;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AKKMKRRKA 9  
 |||||  
 Db 1 AKKMKRRKA 9

RESULT 2  
 ABB49239 standard; Protein; 69 AA.

XX ABB49239;

XX 05-FEB-2002 (first entry)

XX Listeria monocytogenes protein #1943.

XX Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;  
 XX vitamin B12; bacterial infection; disease.

XX Listeria monocytogenes.

XX MO200177335-A2.

XX 18-OCT-2001.

XX 11-APR-2001; 2001WO-FR01118.

XX 11-APR-2000; 2000FR-0004629.

XX

PA (INSP ) INST PASTEUR.

XX Buchrieser C, Frangoul L, Couve E, Rusniok C, Feihi H, Dehoux P;  
 PI Dubsurget O, Chetoui F, Nedjari H, Glaser P, Kunst F, Cossart P;  
 PI Daniels J, Goebel W, Krief J, Kuhn M, Ng E, Vazquez-Boland JH;  
 PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;  
 PI Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L;  
 PI Perez-Diaz J, Aguiro F, Garcia Del Portillo F, Gomez-Lopez N;  
 PI Mahuato E, De Pablo B, Wehland J, Kaerst U, Entian K, Haef J;  
 PI Rose M, Voss H;

XX WPI, 2002-010914/01.

XX Genomic sequence for Listeria monocytogenes, useful e.g. for treatment  
 PT and prevention of Listeria and related bacterial infections, and  
 PT related polypeptides -

XX Claim 6; SEQ ID NO 1944; 192pp; French.

CC The present invention relates to the genome sequence of Listeria  
 CC monocytogenes EGD-e (see ABA03041). The genome sequence and fragments of  
 CC it are useful for selecting probes and primers for detecting genes in L.  
 CC monocytogenes and related organisms, and for studying genetic  
 CC polymorphisms and other genomes. The present invention. Proteins  
 CC encoded by the genome sequence of the present invention. Proteins  
 CC expressed from the genome sequence are useful for raising specific  
 CC antibodies, identification of L. monocytogenes and related organisms,  
 CC for biosynthesis and biodegradation, especially biosynthesis of Vitamin  
 CC B12. The genome sequence and proteins encoded by it are also useful for  
 CC selecting compounds that regulate gene expression and cell replication  
 CC and modulate L. monocytogenes-related diseases. In addition, the genome  
 CC sequence and proteins encoded by it are useful in pharmaceutical and  
 CC vaccines compositions for the treatment or prevention of infections by L.  
 CC monocytogenes and related organisms.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 69 AA;

Query Match 81.4%; Score 35; DB 23; Length 69;  
 Best Local Similarity 87.5%; Pred. No. 34;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 KKKMKRRKA 9  
 |||||  
 Db 60 KKKMKRRKA 67

RESULT 3  
 ABB26198 standard; Protein; 482 AA.

XX ABB26198;

XX 02-JUL-2002 (first entry)

XX Streptococcus polypeptide SEQ ID NO 1572.

XX Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;  
 XX group A streptococcus; Streptococcus pyogenes; antibacterial;  
 XX antiinflammatory; infection; vaccine; meningitis; gene therapy.

XX Streptococcus agalactiae.

XX MO200234771-A2.

XX 02-MAY-2002.

XX 29-OCT-2001; 2001WO-GB04789.

XX 27-OCT-2000; 2000GB-0026333.

XX 24-NOV-2000; 2000GB-0028727.

```

PR 07-MAR-2001; 2001GB-0005640.
XX (CHIR-) CHIRON SPA.
PA (GENO-) INST GENOMIC RES.
XX
PI Telford J, Maignani V, Margarit Ros YI, Grandi G, Fraser C,
PI Tettein H;
XX
XX WPI; 2002-352536/38.
DR N-PSDB; ABN66829.
XX
XX New Streptococcus protein for the treatment or prevention of infection
PT or disease caused by Streptococcus bacteria, such as meningitis, and
PT for detecting a compound that binds to the protein -
XX
XX Claim 1; Page 3308; 4525pp; English.
PS
XX The invention relates to a protein (ABP25413-ABP30895) from group B
CC Streptococcus/GAS (Streptococcus agalactiae) or group A Streptococcus/GAS
CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
CC the specification. The proteins have antibacterial and anti-inflammatory
CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
CC antibodies that bind (I) are used in the manufacture of medicaments for
CC the treatment or prevention of infection or disease caused by
CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
CC Nucleic acids encoding (I) are used to detect Streptococcus in a
CC biological sample. (I) is used to determine whether a compound binds to
CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
CC used as a vaccine or diagnostic composition. The disease caused by
CC Streptococcus that is prevented or treated may be meningitis. Nucleic
CC acid encoding (I) may be used to recombinantly produce (I) and may be
CC used in gene therapy. Antibodies to (I) are used for affinity
CC chromatography, immunoassays, and distinguishing/identifying
CC Streptococcus proteins.
XX
SQ Sequence 482 AA;
XX
Query Match 79.1%; Score 34; DB 23; Length 482;
Best Local Similarity 77.8%; Pred. No. 2.8e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
XX
QY 1 AKMKRRKA 9
Db 467 AKMKRRNA 475
XX
RESULT 4
ABBS7857
ID ABB57857 standard; Protein; 567 AA.
XX
AC ABB57857;
XX
XX 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster polypeptide SEQ ID NO 363.
DE
XX Drosophila; developmental biology; cell signalling; insecticide;
KM pharmaceutical.
XX
XX Drosophila melanogaster.
OS
XX
XX WO200171042-A2.
PN
XX
XX 27-SEP-2001.
PD
XX
XX 23-MAR-2001; 2001WO-US09231.
PF
XX
XX 23-MAR-2000; 2000US-191637P.
PR
XX 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE) PE CORP NY.
PA
XX
XX Venter JC, Adams M, Li FMD, Myers BW,
PI

```

```

XX
XX WPI; 2001-656860/75.
DR N-PSDB; ABL01960.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
XX Disclosure; SEQ ID NO 363; 21pp + Sequence Listing; English.
PS
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABBS7737-ABBS72072).
CC
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 567 AA;
XX
Query Match 79.1%; Score 34; DB 22; Length 567;
Best Local Similarity 77.8%; Pred. No. 3.2e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
XX
QY 1 AKMKRRKA 9
Db 478 AKMKRRTA 486
XX
RESULT 5
AAO09016
ID AAO09016 standard; Protein; 119 AA.
XX
AC AAO09016;
XX
XX 06-NOV-2001 (first entry)
XX
XX Human polypeptide SEQ ID NO 22908.
DE
XX
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KM vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KM tissue growth factor; immunomodulatory; cancer; leukaemia;
KM nervous system disorders; arthritis; inflammation.
XX
XX Homo sapiens.
OS
XX
XX WO200164835-A2.
PN
XX
XX 07-SEP-2001.
PD
XX
XX 26-FEB-2001; 2001WO-US04927.
PF
XX
XX 28-FEB-2000; 2000US-0515126.
PR
XX 18-MAY-2000; 2000US-0577409.
XX
XX (HYSE-) HYSEQ INC.
PA
XX
XX Tang YT, Liu C, Drmanac RT;
PI
XX
XX WPI; 2001-514838/56.
DR N-PSDB; AAI88947.
XX
XX Isolated nucleic acids and polypeptides, useful for preventing
PT diagnosing and treating e.g. leukaemia, inflammation and immune
PT disorders -
XX
XX Claim 20; SEQ ID NO 22908; 1399pp + Sequence Listing; English.
PS
XX
XX The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC

```

CC the encoded proteins (AA000010-AA013910) that exhibit activity elating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activin/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 119 AA;  
  
Query Match 76.7%; Score 33; DB 22; Length 119;  
Best Local Similarity 87.5%; Pred. No. 1.2e+02;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
QY 2 KKKKKRKA 9  
Db 34 KKKKKRKA 41  
  
RESULT 6  
AAG00681  
ID AAG00681 standard; Protein; 120 AA.  
XX  
AC AAG00681;  
XX  
DT 06-OCT-2000 (first entry)  
XX  
DE Human secreted protein, SEQ ID NO: 4762.  
XX  
KW Human, 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
KM gene therapy; chromosome mapping.  
XX  
OS Homo sapiens.  
XX  
PN EPI033401-A2.  
PD 06-SEP-2000.  
XX  
PF 21-FEB-2000; 2000EP-0200610.  
XX  
PR 26-FEB-1999; 99US-0122487.  
XX  
PS (GEST ) GENSET.  
XX  
PI Dumas Milne Edwards J, Duclert A, Giordano J;  
XX  
DR WPI; 2000-500381/45.  
DR N-PSDB; AAC00687.  
XX  
XX  
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for  
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -  
XX  
XX  
XX Claim 13; SEQ ID 4762; 71bp + CD-ROM; English.  
XX  
CC The present sequence is a polypeptide encoded by one of a large number  
CC of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs  
CC were prepared from total human RNAs or polyA+ RNAs derived from 30  
CC different tissues. EST sequences usually correspond mainly to the 3'  
CC untranslated region (UTR) of the mRNA because they are often obtained  
CC from oligo-dT primed cDNA libraries. Such ESTs are not well suited for  
CC isolating cDNA sequences derived from the 5' ends of mRNAs and even in  
CC those cases where longer cDNA sequences have been obtained, the full 5'  
CC UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'  
CC ends and can therefore be used to obtain full length cDNAs and genomic  
CC DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and  
CC chromosome mapping procedures. They are used to obtain upstream

CC regulatory sequences and to design expression and secretion vectors.  
XX  
SQ Sequence 120 AA;  
  
Query Match 76.7%; Score 33; DB 21; Length 120;  
Best Local Similarity 66.7%; Pred. No. 1.2e+02;  
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 AKKKRRKA 9  
Db 30 AKKKRRKA 38  
  
RESULT 7  
AAG00682  
ID AAG00682 standard; Protein; 120 AA.  
XX  
AC AAG00682;  
XX  
DT 06-OCT-2000 (first entry)  
XX  
DE Human secreted protein, SEQ ID NO: 4763.  
XX  
KW Human, 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
KM gene therapy; chromosome mapping.  
XX  
OS Homo sapiens.  
XX  
PN EPI033401-A2.  
PD 06-SEP-2000.  
XX  
PF 21-FEB-2000; 2000EP-0200610.  
XX  
PR 26-FEB-1999; 99US-0122487.  
XX  
PS (GEST ) GENSET.  
XX  
PI Dumas Milne Edwards J, Duclert A, Giordano J;  
XX  
DR WPI; 2000-500381/45.  
DR N-PSDB; AAC00688.  
XX  
XX  
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for  
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -  
XX  
XX  
XX Claim 13; SEQ ID 4763; 71bp + CD-ROM; English.  
XX  
CC The present sequence is a polypeptide encoded by one of a large number  
CC of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs  
CC were prepared from total human RNAs or polyA+ RNAs derived from 30  
CC different tissues. EST sequences usually correspond mainly to the 3'  
CC untranslated region (UTR) of the mRNA because they are often obtained  
CC from oligo-dT primed cDNA libraries. Such ESTs are not well suited for  
CC isolating cDNA sequences derived from the 5' ends of mRNAs and even in  
CC those cases where longer cDNA sequences have been obtained, the full 5'  
CC UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'  
CC ends and can therefore be used to obtain full length cDNAs and genomic  
CC DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and  
CC chromosome mapping procedures. They are used to obtain upstream  
CC regulatory sequences and to design expression and secretion vectors.  
XX  
SQ Sequence 120 AA;  
  
Query Match 76.7%; Score 33; DB 21; Length 120;  
Best Local Similarity 66.7%; Pred. No. 1.2e+02;  
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 AKKKRRKA 9  
Db 30 AKKKRRKA 38

RESULT 8  
 ABG23003  
 ID ABG23003 standard; Protein; 193 AA.  
 AC ABG23003;  
 XX  
 XX 18-FEB-2002 (first entry)  
 DT  
 DE Novel human diagnostic protein #22994.  
 XX  
 XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KM food supplement; medical imaging; diagnostic; genetic disorder.  
 XX  
 OS Homo sapiens.  
 PN WO200175067-A2.  
 XX  
 PD 11-OCT-2001.  
 XX  
 PF 30-MAR-2001; 2001WO-US08631.  
 XX  
 PR 31-MAR-2000; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Drmanac RT, Liu C, Tang YT;  
 DR WPI; 2001-639362/73.  
 XX N-PSDB; AAS67190.  
 XX  
 PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostic, forensic, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -  
 XX  
 PS Claim 20; SEQ ID No 53362; 103pp; English.  
 XX  
 XX The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 XX  
 SQ Sequence 193 AA;  
 Query Match 76.7%; Score 33; DB 22; Length 193;  
 Best Local Similarity 75.6%; Pred. No. 1.8e+02;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 KKMGRKA 9  
 Db 56 KKKMKKA 63

AAU69552  
 ID AAU69552 standard; Protein; 218 AA.  
 XX  
 AC AAU69552;  
 XX  
 XX 30-JAN-2002 (first entry)  
 DT  
 DE Human G protein-coupled receptor from cDNA Seq-2627.  
 XX  
 XX Human; G protein-coupled receptor; nGPR; cancer; tumour;  
 KM thyroid disorder; myxoedema; renal failure; inflammatory condition;  
 KM Crohn's disease; cell differentiation disease; homeostasis disease;  
 KM rheumatoid arthritis; autoimmune disorder; movement disorder;  
 KM central nervous system disorder; stroke; Huntington's disease;  
 KM Tourette's syndrome; Parkinson's disease; Alzheimer's disease;  
 KM viral infection; HIV-1; HIV-2; human immunodeficiency virus;  
 KM metabolic disease; cardiovascular disease; type 2 diabetes; obesity;  
 KM hypotension; hypertension; thrombosis; myocardial infarction;  
 KM atherosclerosis; proliferative disease; hyperproliferative disorder;  
 KM psoriasis; hormonal disorder; polycystic ovarian syndrome;  
 KM alopecia; sexual dysfunction; gene therapy.  
 XX  
 OS Homo sapiens.  
 PN WO200177330-A2.  
 XX  
 PD 18-OCT-2001.  
 XX  
 PF 06-APR-2001; 2001WO-US11330.  
 XX  
 PR 06-APR-2000; 2000US-195093P.  
 PR 06-APR-2000; 2000US-195098P.  
 PR 06-APR-2000; 2000US-195099P.  
 PR 06-APR-2000; 2000US-195148P.  
 PR 06-APR-2000; 2000US-195150P.  
 PR 06-APR-2000; 2000US-195151P.  
 PR 05-SEP-2000; 2000US-230149P.  
 XX  
 PA (PHMA) PHARMACIA & UPJOHN CO.  
 XX  
 PI Vogel G;  
 DR WPI; 2002-010912/01.  
 XX N-PSDB; AAS62884.  
 DR  
 XX  
 PT Novel isolated nucleic acid molecule encoding G protein-coupled  
 PT receptor polypeptide, nGPR-x, useful for treating cancer, Crohn's  
 PT disease, rheumatoid arthritis, Alzheimer's disease, stroke, thrombosis,  
 PT psoriasis -  
 XX  
 PS Claim 31; Page 74; 189pp; English.  
 XX  
 XX The invention relates to an isolated nucleic acid molecule comprising a  
 CC nucleotide sequence that encodes a G protein-coupled receptor  
 CC polypeptide, nGPR-x, vectors and transformed cell expressing the  
 CC protein, antibodies raised against the protein, modulators of the  
 CC protein's activity and methods of isolating the modulators. The n-GPR is  
 CC for purifying a G protein from a sample containing nGPR-x. The n-GPR is  
 CC nucleic acid is useful for recombinantly expressing nGPR-x receptor, for  
 CC detecting expression of the receptor, in the design of antisense and  
 CC other molecules for the suppression of the expression of nGPR-x in a  
 CC cultured cell, tissue or an animal, for therapeutic purposes and to  
 CC provide a model for diseases of conditions characterised by aberrant  
 CC nGPR-x expression. It is also useful in hybridisation assays to detect  
 CC the capacity of cells to express nGPR-x, in diagnosis, to identify  
 CC homologues of nGPR-x in other animals, for screening for restriction  
 CC fragment length polymorphism (RFLP) associated with certain disorders,  
 CC as well as for genetic mapping. The nucleic acid is useful in gene  
 CC therapy. The nGPR is useful as a research tool for identification,  
 CC characterisation and purification of interacting, regulatory proteins.  
 CC The antibody is useful for therapeutic and diagnostic purposes, and in  
 CC purification of nGPR-x. The proteins, nucleic acids and antibodies are  
 CC useful to treat or prevent unregulated cellular growth, such as cancer

RESULT 9

CC cell and tumour growth, and for treating thyroid disorders (e.g.,  
 CC amyloidosis), renal failure, inflammatory conditions (e.g., Crohn's  
 CC disease), diseases related to cell differentiation and homeostasis,  
 CC rheumatoid arthritis, autoimmune disorders, movement disorders, central  
 CC nervous system disorders (e.g., stroke, Huntington's disease, Tourette's  
 CC syndrome, Parkinson's disease, Alzheimer's disease), infections, such as  
 CC viral infections caused by HIV-1 or HIV-2 (human immunodeficiency  
 CC virus, metabolic and cardiovascular diseases and disorders (e.g., type  
 CC 2 diabetes, obesity, hypertension, hyperextension, thrombosis, myocardial  
 CC infarction, atherosclerosis), proliferative diseases and cancers,  
 CC hyperproliferative disorders (such as psoriasis), hormonal disorders  
 CC (e.g., polycystic ovarian syndrome, alopecia), and sexual dysfunction.  
 CC The present sequence represents an n-PCR of the invention.

XX  
 XX  
 SQ Sequence 218 AA;

Query Match 76.7%; Score 33; DB 23; Length 218;  
 Best Local Similarity 75.0%; Pred. No. 2.1e+02;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Oy 2 KKKRRKA 9  
 Db 53 KKKRRRA 60

RESULT 10  
 ABP51251  
 ID ABP51251 standard; Protein; 225 AA.

XX  
 AC ABP51251,  
 XX  
 DT 03-SEP-2002 (first entry)  
 XX  
 DE Human MDDT SEQ ID NO 273.

XX  
 XX Human; MDDT; disease detection and treatment molecule polynucleotide;  
 KM proliferative disorder; hepatitis; psoriasis; cancer; AIDS;  
 KM autoimmune disorder; inflammatory disorder; allergy; multiple sclerosis;  
 KM rheumatoid arthritis; transgenic; gene therapy; antitumor sclerotic;  
 KM hepatotropic; antiinflammatory; antipsoriatic; cyclostatic; anti-HIV;  
 KM antiallergic; antianaemic; antiaethmatic; antiatherosclerotic; anti-gout;  
 KM neuroprotective; antineumatic; antiarthritic.

XX  
 OS Homo sapiens.

XX  
 PN MO200240715-A2.

XX  
 PD 23-MAY-2002.

XX  
 PF 06-SEP-2001; 2001WO-US27628.

XX  
 PR 06-SEP-2000; 2000US-230505P.  
 PR 06-SEP-2000; 2000US-230514P.  
 PR 06-SEP-2000; 2000US-230515P.  
 PR 06-SEP-2000; 2000US-230517P.  
 PR 06-SEP-2000; 2000US-230518P.  
 PR 06-SEP-2000; 2000US-230519P.  
 PR 06-SEP-2000; 2000US-230595P.  
 PR 06-SEP-2000; 2000US-230597P.  
 PR 06-SEP-2000; 2000US-230598P.  
 PR 06-SEP-2000; 2000US-230599P.  
 PR 06-SEP-2000; 2000US-230610P.  
 PR 06-SEP-2000; 2000US-230655P.  
 PR 06-SEP-2000; 2000US-230888P.  
 PR 06-SEP-2000; 2000US-230989P.  
 PR 07-SEP-2000; 2000US-230951P.  
 PR 07-SEP-2000; 2000US-231163P.  
 PR 07-SEP-2000; 2000US-231167P.

XX  
 PA (INCY-) INCYTE GENOMICS INC.

PI Jackson S, Lincoln SE, Altus CM, Dufour GE, Chalup MS, Hillman JL,  
 PI Jones AL, Yu JY, Wright RJ, Gietzen D, Liu TF, Yap PE, Dahl CR,

PI Momiyama MG, Bradley DL, Rohatgi SD, Harris B, Roseberry AM,  
 PI Gershin EH, Peralta CH, David MH, Panzer SR, Flores V, Dafo A;  
 PI Marwaha R, Chen AJ, Chang SC, Au AP, Inman RR;  
 XX WPI; 2002-527544/56.  
 DR N-PSDB; ABQ72469.

XX  
 PT Novel human disease detection and treatment polypeptide, useful in  
 PT diagnosis, prevention or treatment of cell proliferative disorders e.g.  
 PT arteriosclerosis, cirrhosis and an autoimmune/inflammatory disorder  
 PT e.g. AIDS -

XX  
 PS Claim 14; Page 455; 618pp; English.

XX  
 CC The invention relates to an isolated human disease detection and  
 CC treatment (MDDT) polypeptide (I) selected from a polypeptide having a  
 CC sequence selected from 255 sequences (ABP51231-ABP51484) given in the  
 CC specification, a naturally occurring polypeptide comprising a sequence  
 CC having at least 90% identity to (I) or a biologically active or  
 CC immunogenic fragment of (I). (I) is useful for screening a compound for  
 CC effectiveness as an agonist or antagonist, for screening a compound that  
 CC specifically binds (II) or modulates the activity of (I), and for  
 CC preparing a polyclonal or monoclonal antibody by hybridoma technology.  
 CC Nucleic acids (II) (ABQ72449-ABQ72700) encoding (I) are useful for  
 CC screening a compound for effectiveness in altering expression of a target  
 CC polynucleotide comprising. Oligonucleotides and antibodies are useful for  
 CC detecting MDDT in a sample or for assessing toxicity of a test compound,  
 CC in a diagnostic test for a condition or a disease associated with the  
 CC expression of MDDT in a biological sample, for detecting (I) in a sample,  
 CC and for purifying (I) from a sample. A composition comprising (I), an  
 CC agonist or antagonist is useful for treating a disease or condition  
 CC associated with decreased or increased expression of functional MDDT.  
 CC (I) or (II) are useful for diagnosing, treating or preventing disorders  
 CC associated with aberrant expression of MDDT, where the disorders are  
 CC selected from a cell proliferative disorder such as arteriosclerosis,  
 CC cirrhosis, hepatitis, psoriasis, and cancer and AIDS. Addison's disease,  
 CC autoimmune/inflammatory disorder such as AIDS, Addison's disease,  
 CC allergy, anemia, asthma, atherosclerosis, gout, multiple sclerosis or  
 CC rheumatoid arthritis. (II) are useful for creating knockin humanised  
 CC animals or transgenic animals to model human diseases, in somatic or  
 CC gemline gene therapy, to generate a transcrip image of a tissue or cell  
 CC type, for detecting differences in the chromosomal location due to  
 CC translocation or inversion among normal, carrier or affected individuals  
 CC and as hybridisation probes for mapping naturally occurring genomic  
 CC sequences.

XX  
 SQ Sequence 225 AA;

Query Match 76.7%; Score 33; DB 23; Length 225;  
 Best Local Similarity 66.7%; Pred. No. 2.1e+02;  
 Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Oy 1 AKKKRRKA 9  
 Db 37 AKKRRRA 45

RESULT 11  
 AAY49903  
 ID AAY49903 standard; Protein; 505 AA.

XX  
 AC AAY49903;  
 XX  
 DT 26-JAN-2000 (first entry)

XX  
 DE Mouse ROSA26 antisense region protein.

XX  
 DE PCR; genotyping; transgenic animal; Cre recombinase; deleter construct;  
 KM reporter; mouse; expression; primer; genetic engineering.

XX  
 OS Mus sp.

XX  
 PN MO9953017-A2.



```

PD XX 21-OCT-1999.
PP XX 14-APR-1999; 99WO-US08154.
PR XX 15-APR-1998; 98US-0081894.
PA (HUTC-) HUTCHINSON CANCER RES CENT FRED.
XX PI Soriano P, Robertson EJ;
XX DR WPI; 1999-633826/54.
XX N-RSDB; AAZ32469.
PT Vector constructs for making transgenic animals -
XX Example 1; Page 34; 83pp; English.
CC The present invention describes the use of a construct with a sequence
CC homologous to an endogenous, ubiquitously expressed gene locus, to make
CC a transgenic non-human animal which ubiquitously expresses a
CC heterologous DNA. The method for making a genetically engineered
CC non-human animal which ubiquitously expresses a heterologous DNA
CC segment, comprising: (a) introducing into a pluripotent cell a DNA
CC construct, comprising a heterologous DNA segment and at least 100 base
CC pairs (bp) that are homologous to a ubiquitously expressed gene (UEG)
CC from the cell; (b) selecting for cells that carry the heterologous DNA
CC under the UEG promoter control; (c) introducing the selected cell into
CC a developing non-human animal embryo; (d) allowing the embryo to develop
CC to term; and (e) identifying at least one offspring that carries the
CC heterologous DNA under the UEG control. Vectors containing DNA
CC homologous to a UEG locus are useful for developing transgenic animals
CC by positioning a gene of interest under the control of a ubiquitously
CC expressed promoter. The animals may be used to confirm the nature of
CC promoters of ubiquitously expressed genes. The present sequence
CC represents the mouse ROSA26 antisense region protein from the present
CC invention.
XX SQ Sequence 505 AA;
QY Query Match 76.7%; Score 33; DB 20; Length 505;
Best Local Similarity 87.5%; Pred. No. 4.3e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0.
Db 144 KKAARRKA 151
    2 KKKRRKA 9
      |||||
      |||||
      |||||
      |||||
      |||||
AAAG48234
ID AAG48234 standard; Protein; 569 AA.
XX AC AAG48234;
XX DT 18-OCT-2000 (first entry)
DE Arabidopsis thaliana protein fragment SEQ ID NO: 60889.
XX XX
XX Protein identification; signal transduction pathway; metabolic pathway;
XX hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX OS Arabidopsis thaliana.
XX EP1033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-0301439.
XX PR 25-FEB-1999; 99US-0121825.
XX RR 05-MAR-1999; 99US-0123180.
```

PR	03-MAR-1999;	99US-01235348
PR	23-MAR-1999;	99US-01257588
PR	25-MAR-1999;	99US-01262664
PR	29-MAR-1999;	99US-01267685
PR	01-APR-1999;	99US-01276462
PR	06-APR-1999;	99US-01282314
PR	08-APR-1999;	99US-0128714
PR	16-APR-1999;	99US-0129645
PR	19-APR-1999;	99US-0130077
PR	21-APR-1999;	99US-0130449
PR	23-APR-1999;	99US-01305621
PR	28-APR-1999;	99US-01313449
PR	30-APR-1999;	99US-0132048
PR	30-APR-1999;	99US-0132407
PR	04-MAY-1999;	99US-01324854
PR	05-MAY-1999;	99US-01324885
PR	06-MAY-1999;	99US-01324886
PR	06-MAY-1999;	99US-01324887
PR	07-MAY-1999;	99US-0132863
PR	11-MAY-1999;	99US-0134256
PR	14-MAY-1999;	99US-01342119
PR	14-MAY-1999;	99US-0134221
PR	14-MAY-1999;	99US-0134270
PR	18-MAY-1999;	99US-0134470
PR	19-MAY-1999;	99US-01347681
PR	20-MAY-1999;	99US-0135124
PR	21-MAY-1999;	99US-01355153
PR	24-MAY-1999;	99US-0135629
PR	25-MAY-1999;	99US-0136021
PR	27-MAY-1999;	99US-01363692
PR	28-MAY-1999;	99US-01367822
PR	01-JUN-1999;	99US-01372222
PR	03-JUN-1999;	99US-0137528
PR	04-JUN-1999;	99US-0137820
PR	07-JUN-1999;	99US-0137724
PR	08-JUN-1999;	99US-0138094
PR	10-JUN-1999;	99US-0138540
PR	10-JUN-1999;	99US-0138847
PR	14-JUN-1999;	99US-0139119
PR	16-JUN-1999;	99US-0139452
PR	17-JUN-1999;	99US-0139463
PR	17-JUN-1999;	99US-0139492
PR	18-JUN-1999;	99US-0139454
PR	18-JUN-1999;	99US-0139455
PR	18-JUN-1999;	99US-0139457
PR	18-JUN-1999;	99US-0139458
PR	18-JUN-1999;	99US-0139460
PR	18-JUN-1999;	99US-0139461
PR	18-JUN-1999;	99US-0139462
PR	18-JUN-1999;	99US-0139463
PR	18-JUN-1999;	99US-0139750
PR	18-JUN-1999;	99US-0139763
PR	21-JUN-1999;	99US-0139817
PR	22-JUN-1999;	99US-0139869
PR	22-JUN-1999;	99US-0140353
PR	23-JUN-1999;	99US-0140354
PR	24-JUN-1999;	99US-0140695
PR	26-JUN-1999;	99US-0140823
PR	29-JUN-1999;	99US-01409821
PR	30-JUN-1999;	99US-0141867
PR	01-JUL-1999;	99US-0141842
PR	01-JUL-1999;	99US-0142154
PR	02-JUL-1999;	99US-0142055
PR	06-JUL-1999;	99US-0142390
PR	08-JUL-1999;	99US-0142803
PR	09-JUL-1999;	99US-0143290
PR	12-JUL-1999;	99US-0142977
PR	14-JUL-1999;	99US-01435342
PR	14-JUL-1999;	99US-0143624

PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 19-JUL-1999; 99US-0144336.  
PR 20-JUL-1999; 99US-0144337.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149920.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 27-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154033.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155133.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.

PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query March 76.7%; Score 33; DB 21; Length 569;  
Best Local Similarity 75.0%; Pred. No. 4.8e+02;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Oy 2 KKKKKRKA 9  
Db 515 KKKKKRKA 522

RESULT 13  
AAG48233  
ID AAG48233 standard; Protein: 691 AA.

AC AAG48233;

DT 18-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 60888.

XX Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
termination sequence.

XX Arabidopsis thaliana.

FN EP1033405-A2.

PD 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

PR 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126254.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142820.  
PR 12-JUL-1999; 99US-0142877.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 05-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151338.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.

```
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 14-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161982.
PR 28-OCT-1999; 99US-0161993.
PR 28-OCT-1999; 99US-0162142.

Query Match 76.7%; Score 33; DB 21; Length 691;
Best Local Similarity 75.0%; Pred.No. 5.7e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Oy 2 KKKRRRKA 9
Db 637 KKLKKRKA 644

RESULT 14
AAG48232
ID AAG48232 standard; Protein; 705 AA.
XX
AC AAG48232;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 60887.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
```

```
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 14-MAY-1999; 99US-0134376.
PR 18-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135533.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 14-JUN-1999; 99US-0138847.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 24-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140595.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 06-JUL-1999; 99US-0142055.
PR 08-JUL-1999; 99US-0142390.
PR 09-JUL-1999; 99US-0142803.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
```

PR 21-JUL-1999; 99US-0145088.  
 PR 22-JUL-1999; 99US-0145085.  
 PR 22-JUL-1999; 99US-0145087.  
 PR 22-JUL-1999; 99US-0145089.  
 PR 22-JUL-1999; 99US-0145192.  
 PR 23-JUL-1999; 99US-0145145.  
 PR 23-JUL-1999; 99US-0145218.  
 PR 26-JUL-1999; 99US-0145224.  
 PR 27-JUL-1999; 99US-0145276.  
 PR 27-JUL-1999; 99US-0145913.  
 PR 27-JUL-1999; 99US-0145918.  
 PR 28-JUL-1999; 99US-0145951.  
 PR 02-AUG-1999; 99US-0146386.  
 PR 02-AUG-1999; 99US-0146388.  
 PR 02-AUG-1999; 99US-0146389.  
 PR 03-AUG-1999; 99US-0147038.  
 PR 04-AUG-1999; 99US-0147204.  
 PR 04-AUG-1999; 99US-0147302.  
 PR 05-AUG-1999; 99US-0147192.  
 PR 05-AUG-1999; 99US-0147260.  
 PR 06-AUG-1999; 99US-0147303.  
 PR 06-AUG-1999; 99US-0147416.  
 PR 09-AUG-1999; 99US-0147493.  
 PR 09-AUG-1999; 99US-0147935.  
 PR 10-AUG-1999; 99US-0148171.  
 PR 11-AUG-1999; 99US-0148319.  
 PR 12-AUG-1999; 99US-0148341.  
 PR 13-AUG-1999; 99US-0148565.  
 PR 13-AUG-1999; 99US-0148684.  
 PR 16-AUG-1999; 99US-0149368.  
 PR 17-AUG-1999; 99US-0149175.  
 PR 18-AUG-1999; 99US-0149426.  
 PR 20-AUG-1999; 99US-0149722.  
 PR 20-AUG-1999; 99US-0149723.  
 PR 20-AUG-1999; 99US-0149929.  
 PR 23-AUG-1999; 99US-0149902.  
 PR 23-AUG-1999; 99US-0149930.  
 PR 25-AUG-1999; 99US-0150566.  
 PR 26-AUG-1999; 99US-0150884.  
 PR 27-AUG-1999; 99US-0151065.  
 PR 27-AUG-1999; 99US-0151066.  
 PR 27-AUG-1999; 99US-0151080.  
 PR 30-AUG-1999; 99US-0151303.  
 PR 31-AUG-1999; 99US-0151438.  
 PR 01-SEP-1999; 99US-0151930.  
 PR 07-SEP-1999; 99US-0152363.  
 PR 10-SEP-1999; 99US-0153070.  
 PR 13-SEP-1999; 99US-0153758.  
 PR 15-SEP-1999; 99US-0154018.  
 PR 16-SEP-1999; 99US-0154039.  
 PR 20-SEP-1999; 99US-0154779.  
 PR 22-SEP-1999; 99US-0155139.  
 PR 22-SEP-1999; 99US-0155486.  
 PR 24-SEP-1999; 99US-0155659.  
 PR 28-SEP-1999; 99US-0156458.  
 PR 29-SEP-1999; 99US-0156596.  
 PR 04-OCT-1999; 99US-0157117.  
 PR 05-OCT-1999; 99US-0157153.  
 PR 06-OCT-1999; 99US-0157865.  
 PR 07-OCT-1999; 99US-0158029.  
 PR 08-OCT-1999; 99US-0158232.  
 PR 12-OCT-1999; 99US-0158369.  
 PR 13-OCT-1999; 99US-0159293.  
 PR 13-OCT-1999; 99US-0159295.  
 PR 14-OCT-1999; 99US-0159329.  
 PR 14-OCT-1999; 99US-0159330.  
 PR 14-OCT-1999; 99US-0159331.  
 PR 14-OCT-1999; 99US-0159637.  
 PR 14-OCT-1999; 99US-0159638.  
 PR 18-OCT-1999; 99US-0159584.  
 PR 21-OCT-1999; 99US-0160741.

PR 21-OCT-1999; 99US-0160767.  
 PR 21-OCT-1999; 99US-0160768.  
 PR 21-OCT-1999; 99US-0160770.  
 PR 21-OCT-1999; 99US-0160814.  
 PR 21-OCT-1999; 99US-0160815.  
 PR 22-OCT-1999; 99US-0160980.  
 PR 22-OCT-1999; 99US-0160981.  
 PR 22-OCT-1999; 99US-0160989.  
 PR 25-OCT-1999; 99US-0161405.  
 PR 25-OCT-1999; 99US-0161406.  
 PR 25-OCT-1999; 99US-0161406.  
 PR 26-OCT-1999; 99US-0161359.  
 PR 26-OCT-1999; 99US-0161360.  
 PR 26-OCT-1999; 99US-0161361.  
 PR 28-OCT-1999; 99US-0161920.  
 PR 28-OCT-1999; 99US-0161922.  
 PR 28-OCT-1999; 99US-0161932.  
 PR 29-OCT-1999; 99US-0161933.  
 PR 29-OCT-1999; 99US-0162142.

Query Match 76.7%; Score 33; DB 21; Length 705;  
 Best Local Similarity 75.0%; Pred. No. 5.8e+02;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 2 KKKKKRKA 9  
 ||:|||||  
 Db 651 KKKKKRKA 658

RESULT 15  
 ABB65059  
 ID ABB65059 standard; Protein; 221 AA.  
 XX  
 AC ABB65059;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Drosophila melanogaster polypeptide SEQ ID NO 21969.  
 XX  
 KW Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical.  
 XX  
 OS Drosophila melanogaster.  
 XX  
 PN WO200171042-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PF 23-MAR-2001; 2001WO-US09231.  
 XX  
 PR 23-MAR-2000; 2000US-191637P.  
 PR 11-JUL-2000; 2000US-0614150.  
 XX  
 PA (PEKE ) PE CORP NY.  
 XX  
 PI Venter JC, Adams M, Li PWD, Myers EW;  
 XX WPI, 2001-656860/75.  
 DR N-PSDB; ABL09162.  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 PT  
 XX  
 PS Disclosure; SEQ ID NO 21969; 21pp + Sequence Listing; English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins  
 CC (ABB57737-ABB72072).

CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX  
 SO Sequence 221 AA;

Query Match 74.4%; Score 32; DB 22; Length 221;  
 Best Local Similarity 77.8%; Pred. No. 3.1e+02;  
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 AKKKKKRKA 9  
 ||| |||  
 Db 206 AKKAKRKA 214

Search completed: January 29, 2003, 13:59:53  
 Job time : 33.5385 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2003, 13:57:01 ; Search time 12.6933 Seconds  
(without alignments)  
68.168 Million cell updates/sec

Title: US-09-807-949a-137

Perfect score: 43

Sequence: 1 AKMKRRKA 9

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	35	81.4	69	2	AB1333
2	35	81.4	284	1	H69232
3	34	79.1	960	1	T37916
4	33	76.7	211	1	JC2368
5	33	76.7	211	2	G23753
6	32	74.4	98	2	S39410
7	32	74.4	286	2	S34655
8	32	74.4	304	2	A95360
9	32	74.4	391	2	E87449
10	32	74.4	421	2	S64097
11	32	74.4	446	2	T49149
12	32	74.4	1267	2	S53058
13	32	74.4	3724	2	T18427
14	31	72.1	61	2	G83958
15	31	72.1	263	2	T10369
16	31	72.1	265	2	D81402
17	31	72.1	273	2	AC1435
18	31	72.1	273	2	AD1077
19	31	72.1	308	2	G98150
20	31	72.1	337	2	T49514
21	31	72.1	360	2	G90269
22	31	72.1	391	2	T30027
23	31	72.1	422	2	T49513
24	31	72.1	527	2	T04659
25	31	72.1	556	1	A55483
26	31	72.1	557	2	A47162
27	31	72.1	592	2	A41268
28	31	72.1	608	2	D87912
29	31	72.1	628	2	A10382

30	31	72.1	734	2	T23647	hypothetical prote
31	31	72.1	806	2	T23648	hypothetical prote
32	31	72.1	809	2	C96552	hypothetical prote
33	31	72.1	1927	2	A59236	embryonic muscle m
34	30	69.8	65	1	R5KT35	ribosomal protein
35	30	69.8	79	2	A84092	hypothetical prote
36	30	69.8	134	2	T18719	hypothetical prote
37	30	69.8	141	1	H1BPX4	probable core prote
38	30	69.8	207	2	B27626	hypothetical prote
39	30	69.8	254	2	T37887	probable electon
40	30	69.8	260	2	T47391	hypothetical prote
41	30	69.8	264	2	D95870	hypothetical prote
42	30	69.8	277	2	D42400	membrane protein M
43	30	69.8	301	2	T26546	hypothetical prote
44	30	69.8	339	2	D89880	conserved hypochet
45	30	69.8	371	2	H97000	phosphoribosylpyro

#### ALIGNMENTS

RESULT 1  
AB1333  
Hypothetical protein lmo2066 [imported] - Listeria monocytogenes (strain EGD-e)  
C/Species: Listeria monocytogenes  
C/Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 27-Nov-2001  
C/Accession: AB1333  
R/Glaer, P.; Frangoul, L.; Buchtiesser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker, J.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Feil, H.; D.; Jones, L.M.; Karst, U.  
Science 294, 849-852, 2001  
A/Authors: Kref, J.; Kunz, M.; Kunst, F.; Kurapkac, G.; Madueno, E.; Maitounam, A.; Mat ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voess, H.; Wehlend, J.  
A/Title: Comparative genomics of Listeria species  
A/Reference number: AB1077; M01D:21537279; PMID:11679665  
A/Accession: AB1333  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-69 <GLA>  
A/Cross-references: GB:NC\_003210; PIDN:CAD00144.1; PID:gl6411536; GSPDB:GN00177  
A/Experimental source: strain EGD-e  
C/Genetics:  
A/Gene: lmo2066

Query Match 81.4%; Score 35; DB 2; Length 69;  
Best local Similarity 87.5%; Pred. No. 9.6;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 KKMKRRKA 9  
Db 60 KKMKRRKA 67

RESULT 2  
H69232  
M1225-related protein MTH992 - Methanobacterium thermoautotrophicum (strain Delta H)  
N/Alternate names: inosine-5'-monophosphate dehydrogenase related protein IX [msnomer]  
C/Species: Methanobacterium thermoautotrophicum  
C/Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 21-Jul-2000  
C/Accession: H69232  
R/Smith, D.R.; Doucette-Stamm, L.A.; Delouhery, C.; Lee, H.; Dubois, J.; Aldredge, T.; I ki, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.  
J. Bacteriol. 179, 7135-7155, 1997  
A/Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: functi  
A/Reference number: A69000; M01D:98037514; PMID:9371463  
A/Accession: H69232  
A/Status: nucleic acid sequence not shown; translation not shown  
A/Molecule type: DNA  
A/Residues: 1-284 <MTH>  
A/Cross-references: GB:AE000872; GB:AE000666; NID:g2622082; PIDN:AA85489.1; PID:g2622093  
A/Experimental source: strain Delta H  
C/Genetics:

A:Gene: MTH992  
A:Start codon: GTG  
C:Superfamily: conserved hypothetical protein MJ1225; CBS homology  
C:Keywords: duplication  
F:73-120/Domain: CBS homology <CBS>

Query Match 81.4%; Score 35; DB 1; Length 284;  
Best Local Similarity 87.5%; Pred. No. 32;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AKKMKRRK 8  
|:|||||  
Db 246 AKKMKRRK 253

## RESULT 3

T37916  
probable heterochromatin protein - fission yeast (Schizosaccharomyces pombe)  
C:Species: Schizosaccharomyces pombe  
C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999  
C:Accession: T37916  
R:Connor, R.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.; Wood, V.  
submitted to the EMBL Data Library, December 1999  
A:Reference number: 221754  
A:Accession: T37916  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-960 <CON>  
A:Cross-references: EMBL:Z68198; PTDN:CAA92382.1; GSPDB:GNO0066; SPDB:SPAC18G6.02c  
A:Experimental source: strain 972h-; cosmid c18G6  
C:Genetics:  
A:Gene: SPDB:SPAC18G6.02c  
A:Map position: 1

Query Match 79.1%; Score 34; DB 2; Length 960;  
Best Local Similarity 87.5%; Pred. No. 1.4e+02;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AKKMKRRK 8  
|:|||||  
Db 92 AKKMKRRK 99

## RESULT 4

JC2368  
ribosomal protein L13, cytosolic [validated] - rat  
C:Species: Rattus norvegicus (Norway rat)  
C>Date: 21-Jul-2000 #sequence\_revision 21-Jul-2000 #text\_change 21-Jul-2000  
C:Accession: JC2368; PC2232; JC2235; PC2157  
R:Olvera, J.; Wool, I.G.  
Biochem. Biophys. Res. Commun. 201, 102-107, 1994.  
A:Title: The primary structure of rat ribosomal protein L13.  
A:Reference number: JC2235; MUID:94256964; PMID:8198561  
A:Accession: JC2368  
A:Molecule type: mRNA  
A:Residues: 1-211 <OLV1>  
A:Cross-references: EMBL:X78327; NID:G510551; PTDN:CAA55130.1; PID:G510552  
A:Accession: PC2232  
A:Molecule type: protein  
A:Residues: 2-14;26-60;100-180 <OLV2>  
A:Experimental source: clone pL13-2.3  
A>Note: the protein is designated as ribosomal protein L13  
C:Superfamily: rat ribosomal protein L13  
C:Keywords: protein biosynthesis; ribosome  
F:2-211/Product: ribosomal protein L13 #status predicted <MAT>

Query Match 76.7%; Score 33; DB 1; Length 211;  
Best Local Similarity 66.7%; Pred. No. 58;  
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 AKKMKRRK 9  
|:|||||  
Db 30 AKKMKRRK 38

## RESULT 5

S23753  
ribosomal protein L13, cytosolic - human  
N/Alternate names: BB1 protein  
C:Species: Homo sapiens (man)  
C>Date: 27-May-1994 #sequence\_revision 10-Nov-1995 #text\_change 13-Aug-1999  
C:Accession: S23753  
R:Adams, S.M.; Helpe, N.R.; Sharp, M.G.F.; Brammar, W.J.; Walker, R.A.; Varley, J.M.  
Hum. Mol. Genet. 1, 91-96, 1992  
A:Title: Isolation and characterization of a novel gene with differential expression in l  
A:Reference number: S23753; MUID:93244791; PMID:1301162  
A:Accession: S23753  
A:Molecule type: mRNA  
A:Residues: 1-211 <ADA>  
A:Cross-references: EMBL:X64707; NID:G29382; PTDN:CAA45963.1; PID:G29383  
C:Superfamily: rat ribosomal protein L13  
C:Keywords: cytosol; protein biosynthesis; ribosome

Query Match 76.7%; Score 33; DB 2; Length 211;  
Best Local Similarity 66.7%; Pred. No. 58;  
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 AKKMKRRK 9  
|:|||||  
Db 30 AKKMKRRK 38

## RESULT 6

S39410  
nosr protein - Paracoccus denitrificans  
C:Species: Paracoccus denitrificans  
C>Date: 07-Oct-1994 #sequence\_revision 01-Dec-1995 #text\_change 08-Oct-1999  
C:Accession: S39410  
R:Hoeren, F.U.; Berks, B.C.; Ferguson, S.J.; McCarthy, J.E.G.  
Eur. J. Biochem. 218, 49-57, 1993  
A:Title: Sequence and expression of the gene encoding the respiratory nitrous-oxide reduc  
A:Reference number: S39409; MUID:94062841; PMID:8243476  
A:Accession: S39410  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-98 <HOE>  
A:Cross-references: EMBL:X74792; NID:G398932; PTDN:CAA52797.1; PID:G398934

Query Match 74.4%; Score 32; DB 2; Length 98;  
Best Local Similarity 75.0%; Pred. No. 46;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 KKKMKRRK 9  
|:|||||  
Db 68 KKKMKRRK 75

## RESULT 7

S34665  
collagen, cuticular - root-knot nematode (Meloidogyne incognita)  
C:Species: Meloidogyne incognita  
C>Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 20-Sep-1999  
C:Accession: S34665  
R:van der Eycken, W.V.; de Almeida Engler, J.; van Montagu, M.; Gheysen, G.  
submitted to the EMBL Data Library, July 1993  
A:Description: Identification and analysis of a cuticular collagen gene from the plant-p  
A:Reference number: S34665  
A:Accession: S34665  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-286 <VAN>  
A:Cross-references: EMBL:Z24734; NID:G395144; PTDN:CAA80860.1; PID:G395145  
C:Superfamily: unassigned collagens

Query Match 74.4%; Score 32; DB 2; Length 286;  
Best Local Similarity 75.0%; Pred. No. 1.2e+02;



Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 KKKKKRKA 9  
|:|:|:|:  
Db 277 KKKKKRKA 284

## RESULT 8

Probable ABC transporter permease Sma1437 [imported] - *Sinorhizobium meliloti* (strain 10 A95360)

C/Species: *Sinorhizobium meliloti*

C/Date: 24-Aug-2001 #sequence\_revision 24-Aug-2001 #text\_change 30-Sep-2001

C/Accession: A95360

R/Barnett, M.J.; Fisher, R.F.; Jones, T.; Komp, C.; Abola, A.P.; Barloy-Hubler, F.; Bowe

proc. Natl. Acad. Sci. U.S.A. 98, 9883-9888, 2001

A/Title: Nucleotide sequence and predicted functions of the entire *Sinorhizobium meliloti*

A/Reference number: A95262; MUID:21396509; PMID:11481432

A/Accession: A95360

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-304 <R>

A/Cross-references: GB:AE006469; PIDN:AAK65443.1; PID:g14523910; GSPDB:GN00165

A/Experimental source: strain 1021, megaplasmid pSymA

R/Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,

pela, D.; Chain, P.; Cowie, A.; Davies, R.W.; Dreano, S.; Federpiet, N.A.; Fisher, R.F.;

Science 293, 668-672, 2001

A/Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,

hebaull, P.; Vandenberg, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.

A/Title: The composite genome of the legume symbiont *Sinorhizobium meliloti*.

A/Reference number: A96039; MUID:21368234; PMID:11474104

A/Contents: annotation

C/Genetics:

A/Genes: Sma1437

A/Genome: plasmid

Query Match 74.4%; Score 32; DB 2; Length 304;

Best Local Similarity 87.5%; Pred. No. 1.2e+02;

Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AKKKKKRK 8  
|:|:|:|:  
Db 209 AKKKKKRK 216

## RESULT 9

conserved hypothetical protein CCI614 [imported] - *Caulobacter crescentus*

C/Species: *Caulobacter crescentus*

C/Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 20-Apr-2001

C/Accession: E87449

R/Nietman, W.C.; Feldblum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eissen, J.; Heidelberg, J.

n, J.; Laub, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Hart, D.H.; Kolot

proc. Natl. Acad. Sci. U.S.A. 98, 4116-4141, 2001

A/Title: Complete Genome Sequence of *Caulobacter crescentus*.

A/Reference number: AB7249; MUID:21173698; PMID:11259647

A/Accession: E87449

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-391 <STO>

A/Cross-references: GB:AE005673; NID:g13423013; PIDN:AAK23593.1; GSPDB:GN00148

C/Genetics:

A/Genes: CCI614

Query Match 74.4%; Score 32; DB 2; Length 391;

Best Local Similarity 66.7%; Pred. No. 1.5e+02;

Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 AKKKKKRKA 9  
|:|:|:|:  
Db 108 AKKKKKRKA 116

## RESULT 10

hypothetical protein YGL090w - yeast (*Saccharomyces cerevisiae*)

N/Alternate names: hypothetical protein G3179

C/Species: *Saccharomyces cerevisiae*

C/Date: 17-May-1996 #sequence\_revision 17-May-1996 #text\_change 19-Apr-2002

C/Accession: S64097

R/Rieger, M.; Mueller-Auer, S.; Bruckner, M.; Schaefer, M.

submitted to the Protein Sequence Database, May 1996

A/Reference number: S64071

A/Accession: S64097

A/Molecule type: DNA

A/Residues: 1-421 <R>

A/Cross-references: EMBL:Z72612; NID:g1322618; PID:g1322619; GSPDB:GN00007; MIPS:YGL090w

A/Experimental source: strain S288C

C/Genetics:

A/Genes: SGD:LIF1; MIPS:YGL090w

A/Cross-references: SGD:S0003058

A/Map position: 7L

C/Superfamily: *Saccharomyces cerevisiae* hypothetical protein YGL090w

Query Match 74.4%; Score 32; DB 2; Length 421;

Best Local Similarity 85.7%; Pred. No. 1.6e+02;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 KKKKKRKK 8  
|:|:|:|:  
Db 264 KKKKKRKK 270

## RESULT 11

hypothetical protein T20N10.10 - *Arabidopsis thaliana*

C/Species: *Arabidopsis thaliana* (mouse-ear cress)

C/Date: 02-Jun-2000 #sequence\_revision 02-Jun-2000 #text\_change 02-Jun-2000

C/Accession: T49149

R/P'Angelo, M.; Vezzi, A.; Modesto, D.; Piazzi, M.; Valle, G.; Mewes, H.W.; Rudd, S.; Le

submitted to the Protein Sequence Database, April 2000

A/Reference number: Z25017

A/Accession: T49149

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-446 <DAN>

A/Cross-references: EMBL:AL353032; GSPDB:GN00061; ATSP:T20N10.10

A/Experimental source: Cultivar Columbia; BAC clone T20N10

C/Genetics:

A/Genes: ATSP:T20N10.10

A/Map position: 3

Query Match 74.4%; Score 32; DB 2; Length 446;

Best Local Similarity 85.7%; Pred. No. 1.7e+02;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 KKKKKRKK 8  
|:|:|:|:  
Db 345 KKKKKRKK 351

## RESULT 12

Probable membrane protein YMR128w - yeast (*Saccharomyces cerevisiae*)

N/Alternate names: hypothetical protein YMR553.04

C/Species: *Saccharomyces cerevisiae*

C/Date: 08-Jul-1995 #sequence\_revision 01-Sep-1995 #text\_change 02-Feb-2001

C/Accession: S53058

R/Baddock, K.; Churcher, C.

submitted to the EMBL Data Library, March 1995

A/Reference number: S53055

A/Accession: S53058

A/Molecule type: DNA

A/Residues: 1-1267 <BAD>

A;Cross-references: EMBL:Z48622; NID:g728663; PIDD:CAA88553.1; PID:g728667; MIPS:YMR1284  
C;Genetics:  
A;Gene: SGD:ECM16  
A;Cross-references: SGD:S0004735; MIPS:YMR1284  
A;Map position: 13R  
C;Keywords: ATP; nucleotide binding; P-loop; transmembrane protein  
F;414-421/Region: nucleotide-binding motif A (P-loop)  
F;512-517/Region: nucleotide-binding motif B  
F;516-519/Region: DBA motif  
F;922-938/Domain: transmembrane #stratus predicted <TMM>  
Query Match 74.4%; Score 32; DB 2; Length 1267;  
Best Local Similarity 85.7%; Pred. No. 4.1e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 2 KKKRRK 8  
DB 71 KKKRRK 77  
RESULT 13  
T18427  
hypothetical protein C0335C - malaria parasite (Plasmodium falciparum)  
C;Species: Plasmodium falciparum  
C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jun-2000  
C;Accession: T18427  
R;Lawson, D.; Bowman, S.; Barrett, B.  
submitted to the EMBL Data Library, August 1997  
A;Reference number: Z18935  
A;Accession: T18427  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-3724 <IAM>  
A;Cross-references: EMBL:Z98547; NID:e1325376; PID:e1325379; PIDD:CAH1104.1  
C;Genetics:  
A;Intons: 307/1; 1545/2  
A;Note: C0335C  
Query Match 74.4%; Score 32; DB 2; Length 3724;  
Best Local Similarity 85.7%; Pred. No. 1e+03;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 2 KKKRRK 8  
DB 2205 KKKRRK 2211  
RESULT 14  
G83958  
hypothetical protein BH2471 [imported] - Bacillus halodurans (strain C-125)  
C;Species: Bacillus halodurans  
C;Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 15-Jun-2001  
C;Accession: G83958  
R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira  
Nucleic Acids Res. 28, 4317-4331, 2000  
A;Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and  
A;Reference number: A83650; MUID:20512582; PMID:11058132  
A;Accession: G83958  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-61 <STO>  
A;Cross-references: GB:AP01515; GB:BA000004; NID:g10174886; PIDD:BA06190.1; GSPDB:GN00  
A;Experimental source: strain C-125  
C;Genetics:  
A;Gene: BH2471  
Query Match 72.1%; Score 31; DB 2; Length 61;  
Best Local Similarity 85.7%; Pred. No. 47;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 2 KKKRRK 8  
DB 48 KKKRRK 54

RESULT 15  
T10369  
late expression factor 5 protein - Orgyia pseudotsugata nuclear polyhedrosis virus  
C;Species: Orgyia pseudotsugata nuclear polyhedrosis virus, OpMPV  
C;Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 21-Jul-2000  
C;Accession: T10369  
R;Ahrens, C.A.; Russell, R.R.; Funk, C.J.; Evans, J.; Harwood, S.; Rohmann, G.F.  
Virology 229, 381-399, 1997  
A;Title: The sequence of the Orgyia pseudotsugata multinucleocapsid nuclear polyhedrosis  
A;Reference number: Z17011; MUID:97271300; PMID:9126251  
A;Accession: T10369  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-263 <AHR>  
A;Cross-references: EMBL:U75930; NID:g2934903; PIDD:AAC59099.1; PID:g1911346  
Query Match 72.1%; Score 31; DB 2; Length 263;  
Best Local Similarity 85.7%; Pred. No. 1.6e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 2 KKKRRK 8  
DB 193 KKKRRK 199  
Search completed: January 29, 2003, 14:03:26  
Job time : 13.6923 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2003, 13:51:40 ; Search time 6.23077 Seconds  
(without alignments)  
59.910 Million cell updates/sec

Title: US-09-807-949a-137  
Perfect score: 43  
Sequence: 1 AKKMRKKA 9

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues  
Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt 40.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	34	79.1	210	RL13_ICTPU	Q90YV5 Ictalurus p
2	34	79.1	960	CHP1_SCHPO	Q10103 schizosacch
3	33	76.7	210	RL13_BRARE	Q90Z10 brachydanio
4	33	76.7	210	RL13_CHICK	P41125 gallus gall
5	33	76.7	210	RL13_CHICK	Q92313 cricetus
6	33	76.7	210	RL13_HUMAN	P26373 homo sapien
7	33	76.7	210	RL13_MOUSE	P41123 rattus norv
8	32	74.4	421	YGJ0_YEAST	P53150 saccharomyc
9	32	74.4	1267	DHR1_YEAST	Q04217 saccharomyc
10	31	72.1	240	SKR1_HUMAN	Q96C19 homo sapien
11	31	72.1	263	LEB5_NIPOV	Q10344 orgyia pseu
12	31	72.1	557	SASB_ANAPL	Q04791 anae platyr
13	31	72.1	592	GBPI_HUMAN	P32455 homo sapien
14	31	72.1	65	RK35_CVAPA	P14810 cyanophora
15	30	69.8	141	PHCO_BPTA	P04538 bacteriopho
16	30	69.8	254	ETFB_SCHPO	Q9UBH2 schizosacch
17	30	69.8	277	MSWG_STRMU	Q00751 streptococc
18	30	69.8	439	OPSD_LOLSU	Q17094 loligo subu
19	30	69.8	452	OPSD_LOLFO	P24630 loligo forb
20	30	69.8	583	SECD_TREPA	Q92833 homo sapien
21	30	69.8	664	PD13_SHEEP	Q02849 ovie aries
22	30	69.8	917	MSH2_MOUSE	P43248 drosophila
23	30	69.8	1032	ALIA_MOUSE	Q9WV27 mus musculu
24	30	69.8	1170	XPG_MOUSE	P35689 mus musculu
25	30	69.8	1234	JMJ_MOUSE	Q62315 mus musculu
26	30	69.8	1266	JMJ_HUMAN	Q92833 homo sapien
27	30	69.8	66	RL35_BORBU	Q51207 borrelia bu
28	29	67.4	119	PL5_RAT	Q63336 rattus norv
29	29	67.4	126	P15_HUMAN	P53999 homo sapien
30	29	67.4	126	P15_MOUSE	P11031 mus musculu
31	29	67.4	141	YFBJ_CLOPA	Q04687 clostridium
32	29	67.4	189	H5_CHICK	P02259 gallus gall
33	29	67.4	189	H5_CHICK	P02259 gallus gall

34	29	67.4	227	1	DA2D_PHYBI	P21850 phyllomedus
35	29	67.4	230	1	RNFE_SALTY	Q8XEX9 salmomeilla
36	29	67.4	233	1	RNFE_YERPE	Q8XED4 yersinia pe
37	29	67.4	242	1	HXA7_COTJA	P24061 coturnix co
38	29	67.4	287	1	TRUB_AQUAE	O66922 aquifex aeo
39	29	67.4	294	1	RBSK_BACSU	P12254 bacillus su
40	29	67.4	319	1	EXOW_RHIME	P33702 rhizobium m
41	29	67.4	319	1	YOAM_BACSU	P45907 bacillus su
42	29	67.4	366	1	FLHF_BACSU	Q01960 bacillus su
43	29	67.4	392	1	TRPB_BUCMH	O44687 buchnera ap
44	29	67.4	408	1	IF13_MOUSE	O35368 mus musculu
45	29	67.4	453	1	SSP1_YEAST	P38789 saccharomyc

## ALIGNMENTS

RESULT 1  
ID RL13\_ICTPU STANDARD; PRT; 210 AA.  
AC Q90YV5;  
DT 15-JUN-2002 (Rel. 41, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE 60S ribosomal protein L13.  
GN RL13.  
OS Ictalurus punctatus (Channel catfish).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Siluriformes;  
OC Ictaluridae; Ictalurus.  
OX NCBI\_TaxID=7998;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Patterson A.P., Karel A., Liu Z.U.;  
RT "Translational machinery of channel catfish: II. Complementary DNA and  
expression of the complete set of 47 60S ribosomal proteins.";  
RT Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: BELONGS TO THE L13B FAMILY OF RIBOSOMAL PROTEINS.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib.ch](mailto:license@sib.ch)).  
-----  
CC EMBL; AF401567; AAK95139.1; -  
CC InterPro; IPR001380; Ribosomal\_L13B.  
DR pfam; PF01294; Ribosomal\_L13B; 1.  
DR ProDom; PD004443; Ribosomal\_L13B; 1.  
DR PROSITE; PS01104; RIBOSOMAL\_L13B; 1.  
KW Ribosomal protein.  
FT INIT MET 0  
SQ SEQUENCE 210 AA; 24254 MW; 5D97B80DD4E64172 CRC64;  
-----  
Query Match 79.1%; Score 34; DB 1; Length 210;  
Best Local Similarity 66.7%; Pred. No. 12;  
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;  
DB 29 AKKMRKKA 37  
QY 1 AKKMRKKA 9  
DB 29 AKKMRKKA 37  
-----  
RESULT 2  
ID CHP1\_SCHPO STANDARD; PRT; 960 AA.  
AC Q10103;  
DT 01-FEB-1996 (Rel. 33, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Chromo domain protein 1.

GN CHP1 OR SPAC18G6.02C.  
 OS Schizosaccharomyces pombe (Fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 OC Schizosaccharomyces.  
 NCBI\_TaxID=4896;  
 RN  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972;  
 RX MEDLINE=21848401; PubMed=11859360;  
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
 RA Sgouros K., Peat N., Hayles S., Baker S., Sasham D., Bowman S.,  
 RA Brookes K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
 RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,  
 RA Gentles S., Goble A., Hamlin N., Harris P., Hidalgo J., Hodgson A.,  
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jorgels K.,  
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean U.,  
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,  
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitch E.,  
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
 RA Skellon J., Simmonds M., Squares R., Squares S., Stevens K.,  
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
 RA Woodward J., Volkart G., Aert R., Robben J., Grymptre B.,  
 RA Wellens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,  
 RA Gabel C., Fuchs M., Fittz C., Holzer E., Moesl D., Hilbert H.,  
 RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,  
 RA Eger P., Zimmermann W., Medler H., Wambut R., Purnelle B.,  
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,  
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Huet S.M.,  
 RA Lucas M., Rochet M., Galliard C., Tallada V.A., Garzon A., Thode G.,  
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
 RA Cernutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
 RA Sipkavacki G.V., Useery D., Barrell B.G., Nurse P.,  
 RT "The genome sequence of Schizosaccharomyces pombe."  
 RL Nucleic Acids Res. 30:871-880(2002).  
 RN  
 RP CHARACTERIZATION.  
 RX MEDLINE=98391747; PubMed=9722643;  
 RA Doe C.L., Wang G., Chow C.-M., Fricker M.D., Singh P.B., Mellor E.J.,  
 RT "The fission yeast chemo domain encoding gene *chp1(+)* is required for  
 RT chromosome segregation and shows a genetic interaction with *alpha-*  
 RT *tubulin*."  
 RL Nucleic Acids Res. 26:4222-4229(1998).  
 RN  
 RP CHARACTERIZATION.  
 RX MEDLINE=20296735; PubMed=1083390;  
 RA Tison G., Verheide-Hansen J.,  
 RT "Four chemo-domain proteins of Schizosaccharomyces pombe  
 RT differentially repress transcription at various chromosomal  
 RT locations."  
 RL Genetics 155:551-556(2000).  
 CC  
 CC -1- FUNCTION: COMPONENT OF THE KINETOCORE WHICH PLAYS A ROLE IN  
 CC STABILIZING MICROTUBULES AND SO ALLOWING ACCURATE CHROMOSOME  
 CC SEGREGATION.  
 CC  
 CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC  
 CC -1- SIMILARITY: CONTAINS 1 CHROMO DOMAIN.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; Z68198; CA92382.1; -.  
 DR HSBP; P23197; IAP0.  
 DR InterPro: IPR000953; Chromo.  
 DR Pfam; PF00385; chromo; 1.  
 DR SMART; SM00298; CHROMO\_1.  
 DR PROSITE; PS00598; CHROMO\_1; 1.  
 DR PROSITE; PS50013; CHROMO\_2; 1.

KW DNA-binding; Nuclear protein.  
 FT DOMAIN 22 74 CHROMO.  
 FT DOMAIN 142 145 POLY-GLU.  
 SQ SEQUENCE 960 AA; 108726 MW; E92C16BDD05957C CRC64;  
 Query Match 79.1%; Score 34; DB 1; Length 960;  
 Best Local Similarity 87.5%; Pred. No. 51;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 AKKKKKRK 8  
 DB 92 AKKKKKRK 99  
 RESULT 3  
 RL13 BRABE STANDARD; PRT; 210 AA.  
 AC Q90210;  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE 60S ribosomal protein L13.  
 GN RPL13.  
 OS Brachydanio rerio (Zebrafish) (Danio rerio).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;  
 OC Cyprinidae; Danio.  
 NCBI\_TaxID=7955;  
 RN  
 RP SEQUENCE FROM N.A.  
 RA Miller V.M., Rebagliati M.R., Paulson H.L.,  
 RT "Molecular cloning of the zebrafish 60S ribosomal protein L13 (Breast  
 RT basic conserved protein 1 homology)."  
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
 CC  
 CC -1- SIMILARITY: BELONGS TO THE L13E FAMILY OF RIBOSOMAL PROTEINS.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AF35081; AAK63073.1; -.  
 DR InterPro: IPR001380; Ribosomal\_L13E.  
 DR Pfam; PF01294; Ribosomal\_L13e; 1.  
 DR PRODOM; PD004443; Ribosomal\_L13e; 1.  
 DR PROSITE; PS01104; RIBOSOMAL\_L13E; 1.  
 KW Ribosomal protein.  
 FT INIT MET 0 BY SIMILARITY.  
 SQ SEQUENCE 210 AA; 24233 MW; F303B5E4B16DD60 CRC64;  
 Query Match 76.7%; Score 33; DB 1; Length 210;  
 Best Local Similarity 66.7%; Pred. No. 18;  
 Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 AKKKKKRK 9  
 DB 29 AKKKKKRK 37  
 RESULT 4  
 RL13 CHICK STANDARD; PRT; 210 AA.  
 AC P41125;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE 60S ribosomal protein L13 (Breast basic conserved protein 1).  
 GN RPL13 OR BEC1.  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lens fibers;
RX MEDLINE=96437509; PubMed=8840185;
RA Sawada K., Agata K., Eguchi G.;
RT "Characterization of terminally differentiated cell state by
RT categorizing cDNA clones derived from chicken lens fibers.";
RL Int. J. Dev. Biol. 40:531-535(1996).
-----
-1- SIMILARITY: BELONGS TO THE L13E FAMILY OF RIBOSOMAL PROTEINS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.1sb-sib.ch/announce/
CC or send an email to license@1sb-sib.ch).
-----
DR EMBL; D26318; BAA05377.1; -
DR InterPro; IPR001380; Ribosomal_L13E.
DR Pfam; PF01294; Ribosomal_L13e; 1.
DR ProDom; PD004443; Ribosomal_L13E; 1.
DR PROSITE; PS01104; RIBOSOMAL_L13E; 1.
KW Ribosomal protein.
FT INIT MET 0 BY SIMILARITY.
SQ SEQUENCE 210 AA; 24264 MW; D8D005B96A8BA9B4 CRC64;

Query Match 76.7%; Score 33; DB 1; Length 210;
Best Local Similarity 66.7%; Pred. No. 18;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AKCMKRRKA 9
Db 29 ARKRRRKA 37

RESULT 5
RL13_CRIGR STANDARD; PRT; 210 AA.
AC Q92413;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE 60S ribosomal protein L13.
GN RPL13.
OS Cricetus griseus (Chinese hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Cricetus.
OX NCBI_TaxID=10029;
RN [1]
RP SEQUENCE FROM N.A.
RA Sasaki Y., Itoh F., Suzuki H., Hinoda Y., Imai K.;
RT "Identification of genes highly expressed in association with G2
RT arrest induced by DNA damage of Chinese hamster ovary cells by
RT differential display.";
RL Submitted (MAY-1998) TO THE EMBL/GenBank/DBJ databases.
-1- SIMILARITY: BELONGS TO THE L13E FAMILY OF RIBOSOMAL PROTEINS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.1sb-sib.ch/announce/
CC or send an email to license@1sb-sib.ch).
-----
DR EMBL; AB014876; BAA34291.1; -
DR InterPro; IPR001380; Ribosomal_L13E.
DR Pfam; PF01294; Ribosomal_L13e; 1.

```

```

DR ProDom; PD004443; Ribosomal_L13E; 1.
DR PROSITE; PS01104; RIBOSOMAL_L13E; 1.
KW Ribosomal protein.
FT INIT MET 0 BY SIMILARITY.
SQ SEQUENCE 210 AA; 24261 MW; 484429F34C565720 CRC64;

Query Match 76.7%; Score 33; DB 1; Length 210;
Best Local Similarity 66.7%; Pred. No. 18;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AKCMKRRKA 9
Db 29 ARKRRRKA 37

RESULT 6
RL13_HUMAN STANDARD; PRT; 210 AA.
ID RL13_HUMAN
AC P26373; Q9BRX0;
DT 01-AUG-1992 (Rel. 23, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE 60S ribosomal protein L13 (Breast basic conserved protein 1).
GN RPL13 OR BBL1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93244791; PubMed=1301162;
RA Adams S.M., Helps N.R., Sharp M.G.F., Brammar W.J., Walker R.A.,
RA Varley J.M.;
RT "Isolation and characterization of a novel gene with differential
RT expression in benign and malignant human breast tumours.";
RL Hum. Mol. Genet. 1:91-96(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain, Cervix, Lung, Lymph, Pancreas, Placenta, and Uterus;
RA Strausberg R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
-1- TISSUE SPECIFICITY: HIGHER LEVELS OF EXPRESSION IN BENIGN BREAST
CC LESIONS THAN IN CARCINOMAS.
-1- SIMILARITY: BELONGS TO THE L13E FAMILY OF RIBOSOMAL PROTEINS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.1sb-sib.ch/announce/
CC or send an email to license@1sb-sib.ch).
-----
DR EMBL; X64707; CAA45963.1; -
DR EMBL; BC004954; AAH04954.1; -
DR EMBL; BC007345; AAH07345.1; -
DR EMBL; BC007863; AAH07863.1; -
DR EMBL; BC007805; AAH07805.1; -
DR EMBL; BC010994; AAH10994.1; -
DR EMBL; BC013078; AAH13078.1; -
DR EMBL; BC014167; AAH14167.1; -
DR EMBL; BC020804; AAH20804.1; -
DR EMBL; BC027463; AAH27463.1; -
DR EMBL; BC027463; AAH27463.1; -
DR PIR; S23753; S23753.
DR Genew; HGNC:10303; RPL13.
DR MIM; 113703; -
DR InterPro; IPR001380; Ribosomal_L13E.
DR Pfam; PF01294; Ribosomal_L13e; 1.
DR ProDom; PD004443; Ribosomal_L13E; 1.
DR PROSITE; PS01104; RIBOSOMAL_L13E; 1.
KW Ribosomal protein.
FT INIT MET 0 BY SIMILARITY.
FT CONFLICT 111 111 A -> T (IN REF. 1).

```

SQ SEQUENCE 210 AA; 24130 MW; CBBB26F143FAF058 CRC64;  
 Query Match 76.7%; Score 33; DB 1; Length 210;  
 Best Local Similarity 66.7%; Pred. No. 18;  
 Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 AKMKRRKA 9  
 Db 29 ARKIRRRKA 37

RESULT 7  
 RL13\_MOUSE STANDARD; PRT; 210 AA.  
 AC P47963; O9DCH1; O9CRZ9;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE 60S ribosomal protein L13 (A52).  
 GN RPL13.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CBA/J;  
 RA Palacios R., Xie X.;  
 RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Embryonic liver, and kidney;  
 RX MEDLINE=21085660; PubMed=11217851;  
 RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
 RA Aizawa K., Iwawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,  
 RA Saito T., Okazaki Y., Gotohori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochwa H.,  
 RA Kiehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
 RA Schirral L.M., Staudt F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
 RA Blake J., Botfield D., Boujona N., Carninci P., de Bonaldo M.F.,  
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kanliya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima U., Mazzarelli U., Mombereis P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seta T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyokawa K., Wang K.H., Weitz C., Whitaker C., Wilming L.,  
 RA Wysshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,  
 RA Hayashizaki Y.;  
 RA "Functional annotation of a full-length mouse cDNA collection.";  
 RT Nature 409:685-690(2001).  
 RL -1- SIMILARITY: BELONGS TO THE L13E FAMILY OF RIBOSOMAL PROTEINS.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: U26917; AAA69923.1; -;  
 DR EMBL: AK002787; BAB2358.1; -;  
 DR EMBL: AK010969; BAB27309.1; -;  
 DR MGD; MGI:105922; Rpl13.  
 DR InterPro: IPR001380; Ribosomal\_L13E.  
 DR Pfam; PF01294; Ribosomal\_L13E\_1.  
 DR ProDom; PD004443; Ribosomal\_L13E; 1.  
 DR PROSITE; PS01104; RIBOSOMAL\_L13E; 1.  
 DR KW Ribosomal protein.  
 FT INIT MET 0 BY SIMILARITY.

FT CONFLICT 39 39 Q -> L (IN REF. 1).  
 FT CONFLICT 189 195 RLFGIRA -> PTLMOSEQ (IN REF. 1).  
 FT CONFLICT 202 210 AEOVEKCK -> SEQCMKRN (IN REF. 1).  
 SQ SEQUENCE 210 AA; 24174 MW; 3795373F51655F2 CRC64;

QY 1 AKMKRRKA 9  
 Db 29 ARKIRRRKA 37

RESULT 8  
 RL13\_RAT STANDARD; PRT; 210 AA.  
 AC P41123;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 01-FEB-1995 (Rel. 31, Last annotation update)  
 DE 60S ribosomal protein L13.  
 GN RPL13.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.  
 RC STRAIN=Sprague-Dawley; TISSUE=Liver;  
 RX MEDLINE=94256964; PubMed=8198561;  
 RA Olivera J., Wool I.G.;  
 RT "The primary structure of rat ribosomal protein L13.";  
 RL Biochem. Biophys. Res. Commun. 201:102-107(1994).  
 RL -1- SIMILARITY: BELONGS TO THE L13E FAMILY OF RIBOSOMAL PROTEINS.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X78327; CA55130.1; -;  
 DR InterPro: IPR001380; Ribosomal\_L13E.  
 DR Pfam; PF01294; Ribosomal\_L13E; 1.  
 DR ProDom; PD004443; Ribosomal\_L13E; 1.  
 DR PROSITE; PS01104; RIBOSOMAL\_L13E; 1.  
 DR KW Ribosomal protein.  
 FT INIT MET 0  
 FT CONFLICT 55 55 R -> S (IN CDNA).  
 SQ SEQUENCE 210 AA; 24178 MW; 379A6737F30C6EB9 CRC64;

Query Match 76.7%; Score 33; DB 1; Length 210;  
 Best Local Similarity 66.7%; Pred. No. 18;  
 Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 AKMKRRKA 9  
 Db 29 ARKIRRRKA 37

RESULT 9  
 YG10\_YEAST STANDARD; PRT; 421 AA.  
 AC P53150;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE Hypothetical 48.3 kDa protein in NBP35-MPAL2 intergenic region.  
 GN YGI090W.  
 OS Saccharomyces cerevisiae (Baker's yeast).

```

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
RN NCBI_TaxId=4932;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=8288c;
RX MEDLINE=97435481; PubMed9290212;
RA Rieger M., Brueckner M., Schaefer M., Mueller-Auer S.;
RT "Sequence analysis of 203 kilobases from Saccharomyces cerevisiae
chromosome VII.";
RL Yeast 13:1077-1090(1997).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: 272612; CAA96796.1; -
CC SGD: S0003058; LIR1.
CC Hypothetical protein.
KW SEQUENCE 421 AA; 48259 MW; 5BEDC72059E996F CRC64;
SQ
Query Match 74.4%; Score 32; DB 1; Length 421;
Best Local Similarity 85.7%; Pred. No. 54;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 2 KKKKKR 8
DB 264 KKKKKR 270
RESULT 10
DHRI_YEAST STANDARD; PRT; 1267 AA.
ID DHRI_YEAST
AC 004217;
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Probable ATP-dependent RNA helicase DHRI (DEAH-box RNA helicase DHRI).
GN DHRI OR ECM16 OR YMR128W OR YMR553.04.
OS Saccharomyces cerevisiae (Baker's Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxId=4932;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=8288c / AB972;
RA Badcock K., Churcher C., Barrett B.G., Rajandream M.A., Walsh S.V.;
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RN CHARACTERIZATION.
RX MEDLINE=20440184; PubMed10982841;
RA Colley A., Beggs J.D., Tollervey D., Lafontaine D.L.J.;
RT "dhrip, a putative DEAH-box RNA helicase, is associated with the box
C+D snRNP U3.";
Mol. Cell. Biol. 20:7238-7246(2000).
CC -----
CC -1- FUNCTION: PROBABLE ATP-BINDING RNA HELICASE. REQUIRED FOR 18S RNA
CC SYNTHESIS. ASSOCIATES WITH THE U3 SNORNP AND MAY PLAY A ROLE IN
CC RESTRUCTURING OF THE PRE-RRNA.
CC -1- SUBCELLULAR LOCATION: Nuclear; nucleolar.
CC -1- SIMILARITY: BELONGS TO THE DEAD BOX HELICASE FAMILY. DEAH
CC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

CC EMBL: 248622; CAA8553.1; -
CC SGD: S0004735; YMR128W.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002464; DEAH_box.
DR InterPro; IPR001650; Helicase_C.
DR Pfam; PF00271; helicase_C; 1.
DR SMART; SM00487; DEXDC; 1.
DR SMART; SM00490; HELIC; 1.
DR PROSITE; PS00690; DEAH_ATP_HELICASE; 1.
KW ATP-binding; RNA-binding; Helicase; RNA processing.
FT NP BIND 414 421
FT SITE 516 519
FT DEAH BOX.
SQ SEQUENCE 1267 AA; 144954 MW; 576DAAE6D934C77 CRC64;
QY 2 KKKKKR 8
DB 71 KKKKKR 77
RESULT 11
SMS1_HUMAN STANDARD; PRT; 240 AA.
ID SMS1_HUMAN
AC 096C19;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Swiprosin 1.
GN SMS1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RN SEQUENCE FROM N.A.
RC TISSUE=Uterus;
RA Strausberg R.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RN IDENTIFICATION.
RX MEDLINE=21648992; PubMed11789997;
RA Vuadens F., Gasparini D., Deon C., Sanchez J.-C., Hochstrasser D.F.,
RA Schneider P., Tissot J.-D.;
RT "Identification of specific proteins in different lymphocyte
RT populations by proteomic tools.";
RL Proteomics 2:105-111(2002).
CC -----
CC -1- TISSUE SPECIFICITY: Found in lymphocytes; in CD8+ cells.
CC -1- SIMILARITY: CONTAINS 2 EF-HAND CALCIUM-BINDING DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: BC014923; AAH14923.1; -
CC InterPro; IPR002048; EF-hand.
CC Pfam; PF00036; ehand; 2.
CC PRODOM; PD000012; EF-hand; 1.
DR PROSITE; PS00018; EF_HAND; FALSE_NEG.
KW Calcium-binding; Repeat.
FT CA BIND 105 116
FT CA BIND 141 152
FT SEQUENCE 240 AA; 26697 MW; 9FB3FBC3007ABFC2 CRC64;
QY 72.1%; Score 31; DB 1; Length 240;
Best Local Similarity 66.7%; Pred. No. 48;

```

Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 AKKKRKA 9  
 Db 222 AEEKKORKA 230

RESULT 12  
 LEFS NPVOP STANDARD; PRT; 263 AA.  
 ID LEFS NPVOP STANDARD; PRT; 263 AA.  
 AC 010344;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE Late expression factor 5.  
 GN LEF-5.  
 OS Orygia pseudotsugata multicapsid polyhedrosis virus (OPNPV).  
 CC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;  
 CC Nucleopolyhedrovirus.  
 CC NCBI\_TaxID=164623;  
 CC  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97271300; PubMed=9126251;  
 RA Ahrens C.H., Russell R.R., Funk C.J., Evans J., Harwood S.,  
 RA Rohmann G.P.;  
 RT "The sequence of the Orygia pseudotsugata multinnucleocapsid nuclear  
 RT polyhedrosis virus genome.";  
 RL Virology 229:381-399(1997).  
 CC -1- FUNCTION: REQUIRED FOR LATE AND VERY LATE GENE EXPRESSION (BY  
 CC SIMILARITY).  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; U75930; AAC59099.1; -.  
 CC TRANSCRIPTION regulation.  
 CC KW SEQUENCE 263 AA; 30323 MW; 3FE59420A6BD81D4 CRC64;  
 CC SQ

Query Match 72.1%; Score 31; DB 1; Length 263;  
 Best Local Similarity 85.7%; Pred. No. 53;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 KKKRKRK 8  
 Db 193 KKKKRRO 199

RESULT 13  
 SASB ANAPL STANDARD; PRT; 557 AA.  
 ID SASB ANAPL STANDARD; PRT; 557 AA.  
 AC 004751;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Fatty acyl-CoA hydrolase precursor, medium chain (EC 3.1.2.14)  
 DE (Thioesterase B).  
 OS Anas platyrhynchos (Domestic duck).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Archosauria; Aves; Neognathae; Anseriiformes; Anatidae; Anas.  
 CC NCBI\_TaxID=8839;  
 CC [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 26-65.  
 RP TISSUE=Uteroglyal gland;  
 RX MEDLINE=9330823; PubMed=8314791;  
 RA Hwang C.-S., Kolatukudy P.E.;  
 RT "Molecular cloning and sequencing of thioesterase B cDNA and  
 RT stimulation of expression of the thioesterase B gene associated with  
 RT hormonal induction of peroxisome proliferation.";

RL J. Biol. Chem. 268:14278-14284(1993).  
 CC -1- FUNCTION: FATTY ACID BIOSYNTHESIS CHAIN TERMINATION AND RELEASE  
 CC OF THE FREE FATTY ACID PRODUCT IS ACHIEVED BY HYDROLYSIS OF THE  
 CC THIO ESTER BY A THIOESTERASE. THIS THIOESTERASE MAY BE ASSOCIATED  
 CC WITH PEROXISOME PROLIFERATION AND MAY PLAY A ROLE IN THE  
 CC PRODUCTION OF 3-HYDROXY FATTY ACID DIESTER PHEROMONES.  
 CC -1- CATALYTIC ACTIVITY: Oleoyl-(acyl-carrier protein) + H(2)O = [acyl]-  
 CC carrier protein] + oleate.  
 CC -1- TISSUE SPECIFICITY: HIGHEST LEVELS IN UROGYAL GLAND, MUCH LOWER  
 CC IN LIVER AND KIDNEY.  
 CC  
 CC -1- SIMILARITY: BELONGS TO THE TYPE-B CARBOXYL ESTERASE/LIPASE FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; L05493; AAA49223.1; -.  
 CC HSSP; P37967; 10E3  
 DR InterPro; IPR002018; CarboxylesteraseB.  
 DR InterPro; IPR000379; Ser\_estrs\_site.  
 DR Pfam; PF00135; Coesterase; 1.  
 DR PROSITE; PS00122; CARBOXYL ESTERASE B\_1; 1.  
 DR PROSITE; PS00941; CARBOXYL ESTERASE B\_2; 1.  
 KW Fatty acid biosynthesis; Hydrolyase; Signal.  
 FT SIGNAL 1 25  
 FT CHAIN 26 557  
 FT DISULFID 93 122  
 FT ACT\_SITE 227 227  
 FT ACT\_SITE 345 345  
 FT ACT\_SITE 460 460  
 FT CARBOHYD 476 476  
 SQ SEQUENCE 557 AA; 61637 MW; 03E35D90A037F6B0 CRC64;  
 CC

Query Match 72.1%; Score 31; DB 1; Length 557;  
 Best Local Similarity 75.0%; Pred. No. 11e+02;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 AKKKRKRK 8  
 Db 528 AKKKRKRK 535

RESULT 14  
 GBPL HUMAN STANDARD; PRT; 592 AA.  
 ID GBPL HUMAN STANDARD; PRT; 592 AA.  
 AC P32455;  
 DT 01-OCT-1993 (Rel. 27, Created)  
 DT 01-OCT-1993 (Rel. 27, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Interferon-induced guanylate-binding protein 1 (guanine nucleotide-  
 DE binding protein 1).  
 GN GBP1.  
 OS Homo sapiens (Human).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 CC NCBI\_TaxID=9606;  
 CC [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91342675; PubMed=1715024;  
 RA Cheng Y.-S.E., Patterson C.E., Staeheli P.;  
 RT "Interferon-induced guanylate-binding proteins lack an N(T)KXD  
 RT consensus motif and bind GMP in addition to GDP and GTP.";  
 RL Mol. Cell. Biol. 11:4717-4725(1991).  
 RL [2]  
 RP SEQUENCE FROM N.A.  
 RP TISSUE=Uterus;  
 RA Straubenzig R.;  
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
 RL [3]



RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).  
 RX MEDLINE=20140138; PubMed=10676968;  
 RA Prakash B., Praefcke G.J.K., Renaut L., Wittinghofer A., Herrmann C.;  
 RT "Structure of human guanylate-binding protein 1 representing a unique  
 RT class of GMP-binding proteins.";  
 RL Nature 403:567-571(2000).  
 CC -1- FUNCTION: BINDS GTP, GDP AND GMP.  
 CC -1- INDUCTION: BY INTERFERON GAMMA DURING MACROPHAGE ACTIVATION.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; M55542; AAA55871.1; -  
 DR EMBL; RC002666; AAH02666.1; -  
 DR PIR; A41268; A41268.  
 DR PDB; 1DG3; 11-FEB-00.  
 DR Genew; HGNC:4182; GBP1.  
 DR MIM; 600411; -  
 DR InterPro; IPR003191; GBP.  
 DR InterPro; IPR001230; Prenyl\_site.  
 DR Pfam; PF02263; GBP; 1.  
 DR Pfam; PF02841; GBP; 1.  
 KW Interferon induction; GTP-binding; Multigene family; 3D-structure.  
 FT NP\_BIND 45 52 GTP.  
 FT NP\_BIND 97 100 GTP.  
 SQ SEQUENCE 592 AA; 67902 MW; FCO5D1B0FB635467 CRC64;  
 Query Match 72.1%; Score 31; DB 1; Length 592;  
 Best Local Similarity 85.7%; Pred. No. 1.2e+02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 3 KMKRRKA 9  
 DB 582 KMKRRKA 588  
 RESULT 15  
 RK35\_CVAPA STANDARD; PRT; 65 AA.  
 AC P14810;  
 DT 01-APR-1990 (Rel. 14, Created)  
 DT 01-APR-1990 (Rel. 14, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Cyanelle 50S ribosomal protein L35.  
 GN RPL35.  
 OS Cyanophora paradoxa.  
 OC Cyanelle.  
 OC Eukaryota; Glaucocystophyceae; Cyanophoraceae; Cyanophora.  
 OX NCBI\_TaxID=2762;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=LB555 / Pringsheim;  
 RX MEDLINE=90092562; PubMed=2403527;  
 RA Bryant D.A., Stirewalt V.L.;  
 RT "The cyanelle genome of Cyanophora paradoxa encodes ribosomal  
 RT proteins not encoded by the chloroplasts genomes of higher plants.";  
 RL FEBS Lett. 259:273-280(1990).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=LB555 / Pringsheim;  
 RA Stirewalt V.L., Michalowski C.B., Loeffelhardt W., Bohnert H.J.,  
 RA Bryant D.A.;  
 RT "Nucleotide sequence of the cyanelle DNA from Cyanophora paradoxa.";  
 RL Plant Mol. Biol. Rep. 13:327-332(1995).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=LB555 / Pringsheim;  
 RA Loeffelhardt W., Stirewalt V.L., Michalowski C.B., Annarella M.,

RA Farley J.V., Schluchter W.M., Chung S., Newmann-Spallart C.,  
 RA Steiner J.M., Jakowitsch J., Bohnert H.J., Bryant D.A.;  
 RT "The complete sequence of the cyanelle genome of Cyanophora paradoxa:  
 RT the genetic complexity of a primitive plastid.";  
 RL (In) Schenk H.E.A., Herrmann R., Jeon K.W., Mueller N.E.,  
 RL Schwemmler W. (eds.);  
 RL Eukaryotism and Symbiosis, pp.40-48, Springer-Verlag, Heidelberg  
 RL (1997).  
 CC -1- SIMILARITY: BELONGS TO THE L35P FAMILY OF RIBOSOMAL PROTEINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; X17063; CAA34907.1; -  
 DR EMBL; U30821; AAA81192.1; -  
 DR PIR; S07070; R5KT35.  
 DR InterPro; IPR001706; Ribosomal\_L35.  
 DR Pfam; PF01632; Ribosomal\_L35P; 1.  
 DR PRINTS; PR00064; RIBOSOMAL\_L35.  
 DR ProDom; PD003417; Ribosomal\_L35; 1.  
 DR PROSITE; PS00936; RIBOSOMAL\_L35; 1.  
 KW Ribosomal protein, Cyanelle.  
 SQ SEQUENCE 65 AA; 7607 MW; D689E2C85CA95BAF CRC64;  
 Query Match 69.8%; Score 30; DB 1; Length 65;  
 Best Local Similarity 75.0%; Pred. No. 21;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 2 KMKRRKA 9  
 DB 20 KMKRRKA 27

Search completed: January 29, 2003, 14:00:27  
 Job time : 7.23077 secs

**THIS PAGE BLANK (uspto)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2003, 13:56:21 : Search time 25.1538 Seconds  
(without alignments)  
73.723 Million cell updates/sec

Title: US-09-807-949a-137  
Perfect score: 43  
Sequence: 1 AKMKRKA 9

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues  
Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

SPTREMBL 21:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phage:\*
- 10: sp\_plant:\*
- 11: sp\_ricent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_virus:\*
- 16: sp\_bacteriaph:\*
- 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	40	93.0	434	2	Q9ZG66
2	37	86.0	435	2	Q52562
3	35	81.4	69	16	Q8Y5J4
4	35	81.4	166	5	Q95XW9
5	35	81.4	284	17	Q27073
6	35	81.4	633	5	Q97305
7	34	79.1	211	13	Q90YV5
8	34	79.1	434	2	Q91486
9	34	79.1	558	5	Q24202
10	34	79.1	567	5	Q9VIB3
11	33	76.7	128	11	Q9D119
12	33	76.7	207	11	Q9CRZ5
13	33	76.7	211	4	Q9BEX0
14	33	76.7	211	11	Q9DCH1
15	33	76.7	211	13	Q90Z10
16	33	76.7	315	11	Q9D3P0

17	33	76.7	504	11	Q99M45	Q99M45 mus musculu
18	33	76.7	505	11	P97770	P97770 mus musculu
19	33	76.7	665	10	Q9FJ37	Q9FJ37 arabidopsis
20	32	74.4	98	2	Q51704	Q51704 paracoccus
21	32	74.4	105	11	Q9DAD9	Q9DAD9 mus musculu
22	32	74.4	221	5	Q9VW77	Q9VW77 drosophila
23	32	74.4	221	5	Q95S24	Q95S24 drosophila
24	32	74.4	297	5	Q9XY05	Q9XY05 ciona savig
25	32	74.4	304	16	Q92Y88	Q92Y88 ritzubium m
26	32	74.4	305	5	Q25467	Q25467 meloidogyne
27	32	74.4	332	17	Q9HKT3	Q9HKT3 thermoplasma
28	32	74.4	391	16	Q9A7V5	Q9A7V5 caulobacter
29	32	74.4	446	10	Q9LXT5	Q9LXT5 arabidopsis
30	32	74.4	509	5	Q9GS27	Q9GS27 hirudo medi
31	32	74.4	576	11	Q9T168	Q9T168 cricetus
32	32	74.4	697	3	Q9UVG5	Q9UVG5 pichia farl
33	32	74.4	724	2	Q9R9K9	Q9R9K9 paracoccus
34	32	74.4	796	5	Q9XZT9	Q9XZT9 drosophila
35	32	74.4	921	5	Q25989	Q25989 plasmodium
36	32	74.4	989	5	Q9GY99	Q9GY99 leishmania
37	32	74.4	1049	5	Q9Y149	Q9Y149 drosophila
38	32	74.4	3724	5	Q77320	Q77320 plasmodium
39	32	74.4	6815	5	Q917U4	Q917U4 drosophila
40	32	74.4	16215	5	Q9NFS3	Q9NFS3 drosophila
41	31	72.1	61	16	Q9KAI9	Q9KAI9 bacillus ha
42	31	72.1	76	12	Q8VB35	Q8VB35 white spot
43	31	72.1	86	13	Q93227	Q93227 agalychnis
44	31	72.1	124	5	Q9N8V5	Q9N8V5 trypanosoma
45	31	72.1	125	12	Q00256	Q00256 oxygia pseu

#### ALIGNMENTS

RESULT 1

ID Q9ZG66 PRELIMINARY; PRT; 434 AA.

AC Q9ZG66;

DT 01-MAY-1999 (TREMBLrel. 10, Created)

DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)

DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)

DE NDP-hexose 3,4-dehydratase homolog.

GN LANQ.

OS Streptomyces cyanogenus.

OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

OC Actinomycetales; Streptomyces; Streptomycetaceae; Streptomyces.

OX NCBI\_TaxID=80860;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=SI36.

RX MEDLINE=9312695; PubMed=933932;

RA Westrich L., Domm S., Faust B., Bedford D., Hopwood D.A.,

RA Bechthold A.;

RT Cloning and characterization of a gene cluster from Streptomyces

RT cyanogenus SI36 probably involved in lanthionin biosynthesis.";

RL FEMS Microbiol. Lett. 170:381-387(1999).

DR EMBL; AF080233; AAD13547.1; -

DR Interpro; IPR00653; DegT\_DmT\_EryC1.

DR Pfam; PF01041; DegT\_DmT\_EryC1; 1.

SQ SEQUENCE 434 AA; 48166 MW; 8EFEE6450F7EBCA8 CRC64;

Query Match 93.0%; Score 40; DB 2; Length 434;  
Best Local Similarity 88.9%; Pred. No. 6.3;  
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AKMKRKA 9  
Db 70 AKMKRKA 78

RESULT 2  
ID 052562 PRELIMINARY; PRT; 435 AA.

AC 052562;  
 DT 01-JUN-1998 (TREMblrel. 06, Created)  
 DT 01-JUN-1998 (TREMblrel. 06, Last sequence update)  
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
 DE DMP-ribose dehydratase.  
 OS Amycolatopsis mediterranei (Nocardia mediterranei).  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Pseudonocardineae; Pseudonocardaceae; Amycolatopsis.  
 NCBI\_Taxid=33910;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=8699;  
 RX MEDLINE=98174059; PubMed=9512878;  
 RA August P.R., Tang L., Yoon Y.J., Ning S., Mueller R., Yu T.W.,  
 RA Taylor M., Hoffmann D., Kim C.G., Zhang X., Hutchinson C.R.,  
 RA Floss H.G.,  
 RT "Biosynthesis of the ansamycin antibiotic rifamycin: deductions from  
 RT the molecular analysis of the rif biosynthetic gene cluster of  
 RT Amycolatopsis mediterranei S699.";  
 RL Chem. Biol. 5:69-79(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=8699;  
 RX MEDLINE=98165773; PubMed=9497318;  
 RA Kim C.G., Yu T.W., Fryhle C.B., Handa S., Floss H.G.,  
 RT "3-Amino-5-hydroxybenzoic acid synthase, the terminal enzyme in the  
 RT formation of the precursor of mcrN units in rifamycin and related  
 RT antibiotics.";  
 RL J. Biol. Chem. 273:6030-6040(1998).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=8699;  
 RX MEDLINE=21201076; PubMed=11278540;  
 RA Yu T.-W., Mueller R., Zhang X., Draeger G., Kim C.-G.,  
 RA Leister E., Floss H.G.,  
 RT "Mutational analysis and reconstituted expression of the biosynthetic  
 RT genes involved in the formation of 3-amino-5-hydroxybenzoic acid, the  
 RT starter unit of rifamycin biosynthesis in Amycolatopsis mediterranei  
 RT S699.";  
 RL J. Biol. Chem. 276:12546-12555(2001).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=8699;  
 RA August P.R., Tang L., Yoon Y.J., Ning S., Mueller R., Hutchinson C.R.,  
 RA Floss H.G.,  
 RT Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL, AF040570; AAC01730.1; -;  
 DR InterPro; IPR000653; Degt\_DmX\_EryC1.  
 DR Pfam; PF01041; Degt\_DmX\_EryC1; 1.  
 SQ SEQUENCE 435 AA; 47466 MW; 00097C683C6FBD87 CRC64;  
 Query Match 86.0%; Score 37; DB 2; Length 435;  
 Best Local Similarity 77.8%; Pred. No. 24;  
 Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 AKKKRRKA 9  
 Db 72 AKKKRRKA 80

RESULT 3  
 08Y5J4 PRELIMINARY; PRT; 69 AA.  
 AC 08Y5J4;  
 DT 01-MAR-2002 (TREMblrel. 20, Created)  
 DT 01-MAR-2002 (TREMblrel. 20, Last sequence update)  
 DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)  
 DE Hypothetical protein lmo2066.  
 GN LMO2066.  
 OS Listeria monocytogenes.  
 OC Bacteria; Firmicutes; Bacillales; Clostridium group; Bacillales;  
 OC Listeriaceae; Listeria.  
 NCBI\_Taxid=1639;  
 Qy 1 AKKKRRKA 9  
 Db 72 AKKKRRKA 80

RESULT 4  
 095XM9 PRELIMINARY; PRT; 166 AA.  
 AC 095XM9;  
 DT 01-DEC-2001 (TREMblrel. 19, Created)  
 DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
 DE Hypothetical 19.3 kDa protein.  
 GN Y54G2A.3.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
 OC Rhabditidae; Pelodierinae; Caenorhabditis.  
 NCBI\_Taxid=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BRISTOL N2;  
 RX MEDLINE=99069613; PubMed=9851916;  
 RA None;  
 RT "Genome sequence of the nematode C. elegans: a platform for  
 RT investigating biology. The C. elegans Sequencing Consortium.";  
 RL Science 282:2012-2018(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BRISTOL N2;  
 RA Wolan C., Zidanic M.,  
 RT "The sequence of C. elegans cosmid Y54G2A.";  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BRISTOL N2;  
 RA Waterston R.,  
 RT "Direct Submision.";  
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL, AC024817; AA688528.1; -;  
 KW Hypothetical protein.  
 SQ SEQUENCE 166 AA; 19335 MW; D253BFD31146D682 CRC64;  
 Query Match 81.4%; Score 35; DB 5; Length 166;  
 Best Local Similarity 87.5%; Pred. No. 26;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 AKKKRRKA 8  
 Db 60 KKIRRRKA 67

Db 75 AKKMKRRK 82

## RESULT 5

ID 027073 PRELIMINARY; PRT; 284 AA.

AC 027073;

DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)

DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)

DE Inosine-5'-monophosphate dehydrogenase related protein IX.

GN MTH92.

OS Methanobacterium thermoautotrophicum.

OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;

OC Methanobacteriaceae; Methanothermobacter.

OX NCBI\_TaxId=187420;

RN (1)

RP SEQUENCE FROM N.A.

RC STRAIN=DELTA H;

RX MEDLINE=98037514; PubMed=9371463;

RA Aldredge T., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,

RA Harrison D., Hoang L., Keagle P., Lamm W., Pochter B., Qiu D.,

RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,

RA Jivani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,

RA McDougall S., Shiner G., Goyal A., Pietrowski S., Church G.M.,

RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;

RT "Complete genome sequence of Methanobacterium thermoautotrophicum

RT delah: functional analysis and comparative genomics."

RL J. Bacteriol. 179:7135-7155(1997).

DR EMBL; AB008072; AAB85489.1;

DR InterPro; IPR000644; CBS\_domain.

DR Pfam; PF00571; CBS; 4.

DR SMART; SM00116; CBS; 4.

DR Complete proteome.

SK SEQUENCE 284 AA; 31804 MW; B8830566EBB784 CRC64;

Query Match 81.4%; Score 35; DB 17; Length 284;

Best Local Similarity 87.5%; Pred. No. 41;

Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AKKMKRRK 8

Db 246 AKKMKRRK 253

## RESULT 6

097305

ID 097305 PRELIMINARY; PRT; 633 AA.

AC 097305;

DT 01-MAY-1999 (TREMBLrel. 10, Created)

DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)

DE Hypothetical 76.2 kDa protein.

GN PFC1055W, MAL37.42.

OS Plasmodium falciparum (isolate 3D7).

OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

OC NCBI\_TaxId=56329;

OX (1)

RP SEQUENCE FROM N.A.

RC STRAIN=3D7;

RX MEDLINE=99376085; PubMed=10448855;

RA Bowman S., Lawson D., Baaham D., Brown D., Chillingworth T.,

RA Churcher C.M., Craig A., Davies R.M., Devlin K., Felwell T.,

RA Gentles S., Gilliam R., Hamlin N., Harris D., Holroyd S., Hornsby T.,

RA Horrocks P., Jagsels K., Jassal B., Kyes S., McLean J., Moulé S.,

RA Mungall K., Murphy T., Oliver K., Quail M.A., Rajandream M.-A.,

RA Rutter S., Skelton J., Squares R., Squares S., Sulston J.E.,

RA Whitehead S., Woodward J.R., Newbold C., Bartell B.G.;

RT "The complete nucleotide sequence of chromosome 3 of Plasmodium

RT falciparum."

RL Nature 400:532-538(1999).

DR EMBL; AL034559; CAB39052.1; -.  
KW Hypothetical protein.  
SQ SEQUENCE 633 AA; 76173 MW; 018E2FE241697A2 CRC64;

Query Match 81.4%; Score 35; DB 5; Length 633;

Best Local Similarity 100.0%; Pred. No. 81;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 KKKMKRRK 8

Db 196 KKKMKRRK 202

## RESULT 7

090YV5

ID 090YV5 PRELIMINARY; PRT; 211 AA.

AC 090YV5;

DT 01-DEC-2001 (TREMBLrel. 19, Created)

DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)

DE Ribosomal protein L13.

OS Ictalurus punctatus (Channel catfish).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Siluriformes;

OC Ictaluridae; Ictalurus.

OX NCBI\_TaxId=7998;

RN (1)

RP SEQUENCE FROM N.A.

RA Patterson A.P., Karsl A., Liu Z.J.;

RT "Translational machinery of channel catfish: II. Complementary DNA and

RT Expression of the Complete Set of 47 60S Ribosomal Proteins."

RL Submitted (Jul-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF401567; AAK95139.1;

DR InterPro; IPR001380; Ribosomal\_L13B.

DR Pfam; PF01294; Ribosomal\_L13B; 1.

DR Prodom; PD004443; Ribosomal\_L13B; 1.

DR PROSITE; PS01104; RIBOSOMAL\_L13B; UNKNOWN 1.

SK SEQUENCE 211 AA; 24386 MW; 4DB35BBBFFA0FF4 CRC64;

Query Match 79.1%; Score 34; DB 13; Length 211;

Best Local Similarity 66.7%; Pred. No. 50;

Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AKKMKRRKA 9

Db 30 AKKMKRRKA 38

## RESULT 8

0914S6

ID 0914S6 PRELIMINARY; PRT; 434 AA.

AC 0914S6;

DT 01-OCT-2000 (TREMBLrel. 15, Created)

DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)

DE NDP-hexose 3,4-dehydratase UrdQ.

GN URDO.

OS Streptomyces fradiae.

OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

OC Actinomycetales; Streptomycetaceae; Streptomycetaceae; Streptomyces.

OX NCBI\_TaxId=1906;

RN (1)

RP SEQUENCE FROM N.A.

RC STRAIN=F2717;

RX MEDLINE=20547803; PubMed=11094336;

RA Hoffmeister D., Ichino K., Doman S., Faust B., Trefzer A.,

RA Rohrer G., Kirchling A., Fischer C., Kunzel E., Bearden D.W.,

RA "The NDP-sugar co-substrate concentration and the enzyme expression

RT level influence the substrate specificity of glycosyltransferases:

RT cloning and characterization of deoxysugar biosynthetic genes of the

RT urdmycin biosynthetic gene cluster."

RL Chem. Biol. 7:821-831(2000).

```

DR EMBL: AF269227; AAF72550.1; -
DR InterPro: IPR000653; Degt Dnrd EryCl.
DR Pfam: PF01041; Degt Dnrd EryCl. 1
SQ SEQUENCE 434 AA; 48084 MW; AAF0C0125377F4BD3 CRC64;

Query Match
Best Local Similarity 79.1%; Score 34; DB 2; Length 434;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AKMKRRKA 9
Db 70 AKKLKRRKA 78

RESULT 9
Q24202 PRELIMINARY; PRT; 558 AA.
ID Q24202
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE Alpha esterase (Fragment).
GN ALPHA-EST8 OR CG1121.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pearygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxId=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Robin C., Medveczky K.M., Russell R.J., Oakeshott J.G.;
RL J. Mol. Evol. 0:0-0(0).
CC -1- SIMILARITY: BELONGS TO THE TYPE-B CARBOXYLESTERASE/LIPASE FAMILY.
DR EMBL: U51052; AAB0151.1; -.
DR HSPB; P37967; I0E3.
DR FLYBae; FBgn0015576; alpha-Est8.
DR InterPro: IPR002018; CarbesteraseB.
DR InterPro: IPR000379; Ser estse_site.
DR Pfam: PF00135; Coesterase; 1.
DR PROSITE; PS00122; CARBOXYLESTERASE_B_1; 1.
DR PROSITE; PS00941; CARBOXYLESTERASE_B_2; 1.
KW Hydrolyase.
FT NON TER
SQ SEQUENCE 558 AA; 63916 MW; BF6F62C8BF212720 CRC64;

Query Match
Best Local Similarity 79.1%; Score 34; DB 5; Length 558;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AKMKRRKA 9
Db 469 AKKLKRRKA 477

RESULT 10
Q9VIB3 PRELIMINARY; PRT; 567 AA.
ID Q9VIB3
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE ALPHA-EST8 protein.
GN ALPHA-EST8 OR CG1121.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pearygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxId=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

```

```

RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Mortan J.R., Vandel M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abdl J.F., Agbayani A., An H.-J., Andrews-Plankoch C., Baldwin D.,
RA Ballew R.M., Baou A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Borchan M.R., Bouck J., Brokerlein P., Brottier P.,
RA Burtles K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laeko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Mostrel A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclob J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Peinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun B.,
RA Sytrakas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weisenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
CC -1- SIMILARITY: BELONGS TO THE TYPE-B CARBOXYLESTERASE/LIPASE FAMILY.
DR EMBL: A8003671; AAF54012.1; -.
DR HSPB; P37967; I0E3.
DR FLYBae; FBgn0015576; alpha-Est8.
DR InterPro: IPR002018; CarbesteraseB.
DR InterPro: IPR000379; Ser estse_site.
DR Pfam: PF00135; Coesterase; 1.
DR PROSITE; PS00122; CARBOXYLESTERASE_B_1; 1.
DR PROSITE; PS00941; CARBOXYLESTERASE_B_2; 1.
KW Hydrolyase.
SQ SEQUENCE 567 AA; 64831 MW; 07D6ACCT2786F14A CRC64;

Query Match
Best Local Similarity 79.1%; Score 34; DB 5; Length 567;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AKMKRRKA 9
Db 478 AKKLKRRKA 486

RESULT 11
Q9D119 PRELIMINARY; PRT; 128 AA.
ID Q9D119
DT 01-JUN-2001 (TRENBLrel. 17, Created)
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
DE 18 days embryo cDNA, RIKEN full-length enriched library,
DE clone:1110006C21, full insert sequence.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.

```

RC STRAIN=C57BL/6J; TISSUE=EMBRYO;  
 RX MEDLINE=21085660; PubMed=11217851;  
 RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Arakawa T., Hara A., Fukunishi Y., Kono H., Adachi S., Fukuda S.,  
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana T.,  
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
 RA Schiml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Balderelli R., Barsh G.,  
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Gustinich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,  
 RA Wyszaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohlsuki S.,  
 RA Hayashizaki Y.,  
 RA "Functional annotation of a full-length mouse cDNA collection.";  
 RT Nature 409:685-690(2001).  
 RL Nature 409:685-690(2001).  
 DB 30 ARKRRKA 38

Query Match 76.7%; Score 33; DB 11; Length 128;  
 Best Local Similarity 66.7%; Pred. No. 51;  
 Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AKKRRKA 9  
 Db 30 ARKRRKA 38

RESULT 12  
 Q9CRZ9 PRELIMINARY; PRT; 207 AA.  
 AC Q9CRZ9;  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
 DE Ribosomal protein L13 (Fragment).  
 GN RPL13.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=EMBRYONIC LIVER;  
 RX MEDLINE=21085660; PubMed=11217851;  
 RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Arakawa T., Hara A., Fukunishi Y., Kono H., Adachi S., Fukuda S.,  
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana T.,  
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
 RA Schiml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Balderelli R., Barsh G.,  
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Gustinich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,  
 RA Wyszaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohlsuki S.,  
 RA Hayashizaki Y.,

RT "Functional annotation of a full-length mouse cDNA collection.";  
 RL Nature 409:685-690(2001).  
 DB 30 ARKRRKA 38

Query Match 76.7%; Score 33; DB 11; Length 207;  
 Best Local Similarity 66.7%; Pred. No. 76;  
 Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AKKRRKA 9  
 Db 30 ARKRRKA 38

RESULT 13  
 Q9BPX0 PRELIMINARY; PRT; 211 AA.  
 AC Q9BPX0;  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Hypothetical 24.3 kDa protein (Ribosomal protein L13) (Unknown)  
 DE (Protein for MGC:115073) (Protein for MGC:117534).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=LUNG;  
 RA Strausberg R.;  
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=LUNG;  
 RA Strausberg R.;  
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=LYMPH, AND UTERUS;  
 RA Strausberg R.;  
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=PLACENTA;  
 RA Strausberg R.;  
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=CERVIX;  
 RA Strausberg R.;  
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=BRIN;  
 RA Strausberg R.;  
 RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.  
 RN [7]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=CERVIX;  
 RA Strausberg R.;  
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL, BC004954; AAH04954.1; -  
 DR EMBL, BC014167; AAH14167.1; -  
 DR EMBL, BC007345; AAH07345.1; -  
 DR EMBL, BC007563; AAH07563.1; -  
 DR EMBL, BC007805; AAH07805.1; -

DR EMBL; BC010994; AAH10994.1; -  
 DR EMBL; BC013078; AAH13078.1; -  
 DR EMBL; BC020804; AAH20804.1; -  
 DR EMBL; BC027463; AAH27463.1; -  
 DR InterPro; IPR001380; Ribosomal\_L13E.  
 DR Pfam; PFO1294; Ribosomal\_L13e; 1.  
 DR ProDom; PD004443; Ribosomal\_L13E; 1.  
 DR PROSITE; PS01104; RIBOSOMAL\_L13E; UNKNOWN\_1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 211 AA; 24261 MW; DB9FCS7768EBDE CRC64;

Query Match 76.7%; Score 33; DB 4; Length 211;  
 Best Local Similarity 66.7%; Pred. No. 77;  
 Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 1 AKKKRRKA 9  
 DB 30 ARKIRRRKA 38

## RESULT 14

ID Q9DCH1 PRELIMINARY; PRT; 211 AA.  
 AC Q9DCH1;  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
 DE Ribosomal protein L13.  
 GN RPL13.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=KIDNEY;  
 RC MEDLINE=21085660; PubMed=11217851;  
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 Arakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,  
 Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,  
 Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
 Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 Pleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
 Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
 Schirai L.M., Straubli F., Suzuki R., Tomita M., Wagner I., Washio T.,  
 Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barch G.,  
 Blake J., Botfield D., Bjunga N., Carninci P., de Bonaldo M.F.,  
 Brownstein M.J., Bult C., Flechner C., Fujita M., Gariboldi M.,  
 Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
 Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
 Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
 Suzuki H., Toyooka K., Wang K.H., Wells C., Whitaker C., Wilming L.,  
 Wysshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohno S.,  
 Hayashizaki Y.;  
 RA "Functional annotation of a full-length mouse cDNA collection.";  
 RT Nature 409:685-690(2001).  
 RL EMBL; AK002787; BAB22358.1; -  
 DR EMBL; AK002787; BAB22358.1; -  
 DR WGD; MGI:105922; Rpl13.  
 DR InterPro; IPR001380; Ribosomal\_L13E.  
 DR Pfam; PFO1294; Ribosomal\_L13e; 1.  
 DR ProDom; PD004443; Ribosomal\_L13E; 1.  
 DR PROSITE; PS01104; RIBOSOMAL\_L13E; UNKNOWN\_1.  
 SQ SEQUENCE 211 AA; 24305 MW; 27B1D4B97A5A1B74 CRC64;

Query Match 76.7%; Score 33; DB 11; Length 211;  
 Best Local Similarity 66.7%; Pred. No. 77;  
 Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 1 AKKKRRKA 9  
 DB 30 ARKIRRRKA 38

RESULT 15  
 ID Q90Z10 PRELIMINARY; PRT; 211 AA.  
 AC Q90Z10;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
 DE 60S ribosomal protein L13.  
 DE Brachydanio rerio (Zebrafish) (Zebra danio).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;  
 OC Cyprinidae; Danio.  
 NX NCBI\_TaxID=7955;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Miller V.M., Rebagliati M.R., Paulson H.L.;  
 RT "Molecular Cloning of the Zebrafish 60S Ribosomal Protein L13 (Breast  
 Basic Conserved Protein 1 Homolog).";  
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF385081; AAK63073.1; -  
 DR InterPro; IPR001380; Ribosomal\_L13E.  
 DR Pfam; PFO1294; Ribosomal\_L13e; 1.  
 DR ProDom; PD004443; Ribosomal\_L13E; 1.  
 DR PROSITE; PS01104; RIBOSOMAL\_L13E; UNKNOWN\_1.  
 KW Ribosomal protein.  
 SQ SEQUENCE 211 AA; 24364 MW; E32706629A5A94E6 CRC64;

Query Match 76.7%; Score 33; DB 13; Length 211;  
 Best Local Similarity 66.7%; Pred. No. 77;  
 Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 1 AKKKRRKA 9  
 DB 30 ARKIRRRKA 38

Search completed: January 29, 2003, 14:02:26  
 Job time : 27.1538 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: January 29, 2003, 13:57:26 ; Search time 11.0769 Seconds  
(without alignments)  
23.906 Million cell updates/sec

Title: US-09-807-949A-137

Perfect score: 43

Sequence: 1 AKKMKRRKA 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Database : Issued Patents AA:\*

1: /cgn2\_6/prodata/1/1aa/5A COMB.pep:\*

2: /cgn2\_6/prodata/1/1aa/5B COMB.pep:\*

3: /cgn2\_6/prodata/1/1aa/6A COMB.pep:\*

4: /cgn2\_6/prodata/1/1aa/6B COMB.pep:\*

5: /cgn2\_6/prodata/1/1aa/PCTUS COMB.pep:\*

6: /cgn2\_6/prodata/1/1aa/backfill1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	33	76.7	505	4	US-09-291-541-10
2	32	74.4	921	2	US-08-568-459A-8
3	32	74.4	921	2	US-08-487-826B-8
4	32	74.4	921	4	US-09-210-288-8
5	31	72.1	280	4	US-09-601-478-5
6	31	72.1	280	4	US-09-601-478-8
7	31	72.1	592	2	US-08-736-770-6
8	30	69.8	384	4	US-09-134-001C-3437
9	30	69.8	543	2	US-08-823-516-143
10	29	67.4	27	1	US-08-231-730A-27
11	29	67.4	27	1	US-08-427-001C-27
12	29	67.4	27	1	US-08-457-171-27
13	29	67.4	27	1	US-08-457-171-27
14	29	67.4	27	2	US-08-505-486-27
15	29	67.4	27	3	US-08-689-489C-27
16	29	67.4	27	3	US-08-601-028-27
17	29	67.4	27	3	US-09-340-154-27
18	29	67.4	27	4	US-09-232-802A-27
19	29	67.4	27	4	US-09-482-611B-27
20	29	67.4	27	5	PCT-US94-06176-27
21	29	67.4	27	5	PCT-US94-12550-27
22	29	67.4	27	5	PCT-US95-04335-27
23	29	67.4	27	5	PCT-US95-04718-27
24	29	67.4	27	5	PCT-US95-09338-27
25	29	67.4	27	5	PCT-US95-09339-27
26	29	67.4	455	4	US-08-840-767-43
27	29	67.4	456	4	US-08-840-767-11

28	29	67.4	976	4	US-08-894-997-50	Sequence 50, Appl
29	28	65.1	13	4	US-09-025-596-103	Sequence 103, App
30	28	65.1	38	4	US-09-461-697-222	Sequence 222, App
31	28	65.1	41	4	US-09-461-697-220	Sequence 220, App
32	28	65.1	47	4	US-09-227-357-251	Sequence 251, App
33	28	65.1	51	4	US-09-461-697-218	Sequence 218, App
34	28	65.1	57	4	US-09-461-697-216	Sequence 216, App
35	28	65.1	62	4	US-09-461-697-212	Sequence 212, App
36	28	65.1	64	4	US-09-461-697-210	Sequence 210, App
37	28	65.1	70	4	US-09-461-697-208	Sequence 208, App
38	28	65.1	76	4	US-09-461-697-206	Sequence 206, App
39	28	65.1	93	4	US-09-461-697-204	Sequence 204, App
40	28	65.1	101	4	US-09-117-257-38	Sequence 38, Appl
41	28	65.1	160	4	US-09-489-352-38	Sequence 38, Appl
42	28	65.1	160	4	US-09-516-914-19	Sequence 19, Appl
43	28	65.1	182	4	US-09-258-257-2	Sequence 2, Appl
44	28	65.1	210	2	US-09-258-371-2	Sequence 2, Appl
45	28	65.1	210	2	US-09-258-371-2	Sequence 2, Appl

#### ALIGNMENTS

```

RESULT 1
; US-09-291-541-10
; Sequence 10, Application US/09291541
; Patent No. 5461864
; GENERAL INFORMATION:
; APPLICANT: Soriano, Philippe
; APPLICANT: Robertson, Elizabeth J.
; TITLE OF INVENTION: METHODS AND VECTOR CONSTRUCTS FOR MAKING TRANSGENIC
; TITLE OF INVENTION: NON-HUMAN ANIMALS WHICH UNBIOGENOUSLY EXPRESS A
; FILE REFERENCE: 145384-44-1
; CURRENT APPLICATION NUMBER: US/09/291,541
; EARLIER FILING DATE: 1999-04-14
; EARLIER APPLICATION NUMBER: US 60/081,894
; EARLIER FILING DATE: 1998-04-15
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 10
; LENGTH: 505
; TYPE: PRT
; ORGANISM: Murine amino acid sequence of ROSA26 antisense region
; US-09-291-541-10

Query Match      76.7%; Score 33; DB 4; Length 505;
Best Local Similarity 87.5%; Pred. No. 1.8e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Cy      2 KKMKRRKA 9
Db      144 KKMKRRKA 151

RESULT 2
; US-08-568-459A-8
; Sequence 8, Application US/08568459A
; Patent No. 5849306
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhaun
; APPLICANT: Wellens, Thomas B.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach

```

STATE: California  
COUNTRY: US  
ZIP: 92660  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/568,459A  
FILING DATE: 07-DEC-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Israelien, Ned  
REGISTRATION NUMBER: 29,655  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 235-8550  
TELEFAX: (619) 235-0176  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 921 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
ORGANISM: Plasmodium falciparum  
US-08-568-459A-8

Query Match 74.4%; Score 32; DB 2; Length 921;  
Best Local Similarity 85.7%; Pred. No. 4.4e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 KKKRRK 8  
|||:|  
Db 912 KKKRRK 918

RESULT 3  
US-08-487-826B-8  
Sequence 8, Application US/08487826B  
Patent No. 593827  
GENERAL INFORMATION:  
APPLICANT: Sim, Kim L.  
APPLICANT: Chitnis, Chetan  
APPLICANT: Miller, Louis H.  
APPLICANT: Peterson, David S.  
APPLICANT: Su, Xin-zhaun  
APPLICANT: Wellens, Thomas E.  
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX  
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS  
NUMBER OF SEQUENCES: 45  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Knobbe Martens Olson & Bear  
STREET: 620 Newport Center Drive 16th Floor  
CITY: Newport Beach  
STATE: California  
COUNTRY: US  
ZIP: 92660  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/487,826B  
FILING DATE: 10-SEP-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Israelien, Ned  
REGISTRATION NUMBER: 29,655

REFERENCE/DOCKET NUMBER: NIH121.001CP1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 235-8550  
TELEFAX: (619) 235-0176  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 921 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
ORGANISM: Plasmodium falciparum  
US-08-487-826B-8

Query Match 74.4%; Score 32; DB 2; Length 921;  
Best Local Similarity 85.7%; Pred. No. 4.4e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 KKKRRK 8  
|||:|  
Db 912 KKKRRK 918

RESULT 4  
US-09-210-288-8  
Sequence 8, Application US/09210288  
Patent No. 6392026  
GENERAL INFORMATION:  
APPLICANT: Sim, Kim L.  
APPLICANT: Chitnis, Chetan  
APPLICANT: Miller, Louis H.  
APPLICANT: Peterson, David S.  
APPLICANT: Su, Xin-zhaun  
APPLICANT: Wellens, Thomas E.  
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX  
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS  
NUMBER OF SEQUENCES: 37  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Knobbe Martens Olson & Bear  
STREET: 620 Newport Center Drive 16th Floor  
CITY: Newport Beach  
STATE: California  
COUNTRY: US  
ZIP: 92660  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/210,288  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Fuller, Michael  
REGISTRATION NUMBER: 36,516  
REFERENCE/DOCKET NUMBER: NIH121.1FWDV1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 235-8550  
TELEFAX: (619) 235-0176  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 921 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
ORGANISM: Plasmodium falciparum  
US-09-210-288-8

Query Match 74.4%; Score 32; DB 4; Length 921;  
Best Local Similarity 85.7%; Pred. No. 4.4e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 KKKKKRKA 8  
|||:||||  
DB 912 KKKKKRKA 918

RESULT 5  
US-09-601-478-5  
; Sequence 5, Application US/09601478  
; Patent No. 6403785  
; GENERAL INFORMATION:  
; APPLICANT: Oetuka Pharmaceutical Co., Ltd.  
; TITLE OF INVENTION: Human TSC403 gene and human INGLL gene  
; FILE REFERENCE: Q60193  
; CURRENT APPLICATION NUMBER: US/09/601,478  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: JP H10-134679  
; PRIOR FILING DATE: 1998-04-28  
; PRIOR APPLICATION NUMBER: JP H10-73234  
; PRIOR FILING DATE: 1998-03-05  
; PRIOR APPLICATION NUMBER: JP H10-38133  
; PRIOR FILING DATE: 1998-02-03  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 5  
; LENGTH: 280  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: human embryonic brain cDNA library  
US-09-601-478-5

Query Match 72.1%; Score 31; DB 4; Length 280;  
Best Local Similarity 77.8%; Pred. No. 2.3e+02;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AKKKRKA 9  
|||:||||  
DB 186 AKKKRKA 194

RESULT 6  
US-09-601-478-8  
; Sequence 8, Application US/09601478  
; Patent No. 6403785  
; GENERAL INFORMATION:  
; APPLICANT: Oetuka Pharmaceutical Co., Ltd.  
; TITLE OF INVENTION: Human TSC403 gene and human INGLL gene  
; FILE REFERENCE: Q60193  
; CURRENT APPLICATION NUMBER: US/09/601,478  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: JP H10-134679  
; PRIOR FILING DATE: 1998-04-28  
; PRIOR APPLICATION NUMBER: JP H10-73234  
; PRIOR FILING DATE: 1998-03-05  
; PRIOR APPLICATION NUMBER: JP H10-38133  
; PRIOR FILING DATE: 1998-02-03  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 8  
; LENGTH: 280  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; OTHER INFORMATION: human embryonic brain cDNA library  
US-09-601-478-8

Query Match 72.1%; Score 31; DB 4; Length 280;  
Best Local Similarity 77.8%; Pred. No. 2.3e+02;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AKKKRKA 9  
|||:||||  
DB 186 AKKKRKA 194

RESULT 7  
US-08-736-770-6  
; Sequence 6, Application US/08736770  
; Patent No. 5871965  
; GENERAL INFORMATION:  
; APPLICANT: Bandman, Olga  
; APPLICANT: Au-Young, Janice  
; APPLICANT: Hillman, Jennifer L.  
; TITLE OF INVENTION: NOVEL HUMAN GUANYLATE BINDING PROTEINS  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: US  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: PASTE0 Version 1.5  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/736,770  
; FILING DATE: Filed Herewith  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0145 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-855-0555  
; TELEFAX: 415-845-4166  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 592 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; IMMEDIATE SOURCE:  
; LIBRARY: GenBank  
; CLONE: 183002  
US-08-736-770-6

Query Match 72.1%; Score 31; DB 2; Length 592;  
Best Local Similarity 85.7%; Pred. No. 4.3e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 KKKRKA 9  
|||:||||  
DB 582 KKKRKA 588

RESULT 8  
US-09-134-001C-3437  
; Sequence 3437, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09/134,001C  
; CURRENT FILING DATE: 1998-08-13

PRIOR APPLICATION NUMBER: US 60/064,964  
PRIOR FILING DATE: 1997-11-08  
PRIOR APPLICATION NUMBER: US 60/055,779  
PRIOR FILING DATE: 1997-08-14  
NUMBER OF SEQ ID NOS: 5674  
SEQ ID NO 3437  
LENGTH: 384  
TYPE: PRT  
ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-3437

Query Match 69.8%; Score 30; DB 4; Length 384;  
Best Local Similarity 66.7%; Pred. No. 4.3e+02;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 AKMKRRKA 9  
DB 325 AKMKRRKA 333

RESULT 9  
US-08-823-516-143  
Sequence 143; Application US/08823516  
Patent No. 5994069  
GENERAL INFORMATION:  
APPLICANT: Hall, Jeff G.  
APPLICANT: Lyamichev, Victor I.  
APPLICANT: Maest, Andreea L.  
APPLICANT: Brow, Mary Ann D.  
TITLE OF INVENTION: Detection Of Nucleic Acids By Multiple  
NUMBER OF SEQUENCES: 163  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Medlen & Carroll, LLP  
STREET: 220 Montgomery Street, Suite 2200  
CITY: San Francisco  
STATE: California  
COUNTRY: United States Of America  
ZIP: 94104

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/823,516  
FILING DATE: 24-MAR-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US97/01072  
FILING DATE: 21-JAN-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/759,038  
FILING DATE: 02-DEC-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/758,314  
FILING DATE: 29-NOV-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/756,386  
FILING DATE: 02-DEC-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/682,853  
FILING DATE: 12-JUL-1996  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/599,491  
FILING DATE: 24-JAN-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Ingolia, Diane E.  
REGISTRATION NUMBER: 40,027  
REFERENCE/DOCKET NUMBER: FORS-02736  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 705-8410  
TELEFAX: (415) 397-8338

INFORMATION FOR SEQ ID NO: 143:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 543 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
TOPOLOGY: not relevant  
MOLECULE TYPE: protein  
US-08-823-516-143

Query Match 69.8%; Score 30; DB 2; Length 543;  
Best Local Similarity 85.7%; Pred. No. 5.8e+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 KMKRRK 8  
DB 534 KMKRRK 540

RESULT 10  
US-08-231-730A-27  
Sequence 27; Application US/08231730A  
Patent No. 5561107  
GENERAL INFORMATION:  
APPLICANT: JAYNES, JESSE M.  
APPLICANT: JULIAN, GORDON R.  
TITLE OF INVENTION: METHOD OF ENHANCING WOUND HEALING BY STIMULATING FIBROBLAST ANI  
NUMBER OF SEQUENCES: 46  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: STEVEN J. HULTQUIST  
ADDRESS: INTELLECTUAL PROPERTY/TECHNOLOGY LAW  
STREET: 200 PARK DRIVE, SUITE 210  
CITY: RESEARCH TRIANGLE PARK  
STATE: NORTH CAROLINA  
COUNTRY: USA  
ZIP: 27709  
COMPUTER READABLE FORM:  
MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.4 MB STORAGE  
COMPUTER: APPLE MACINTOSH  
OPERATING SYSTEM: MACINTOSH  
SOFTWARE: N.S. WORD 5.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/231,730A  
FILING DATE: 04-20-94  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/225,476  
FILING DATE: 04-08-94  
APPLICATION NUMBER: 08/039,620  
FILING DATE: 06-04-93  
APPLICATION NUMBER: 08/148,491  
FILING DATE: 11-08-93  
APPLICATION NUMBER: 08/148,889  
FILING DATE: 11-08-93  
ATTORNEY/AGENT INFORMATION:  
NAME: HULTQUIST, STEVEN J.  
REGISTRATION NUMBER: 28021  
REFERENCE/DOCKET NUMBER: 4013-106  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (919) 990-9531  
TELEFAX: (919) 990-9532  
INFORMATION FOR SEQ ID NO: 27:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 27  
TYPE: AMINO ACID  
TOPOLOGY: LINEAR  
MOLECULE TYPE: PEPTIDE  
DESCRIPTION: NO  
HYPOTHEICAL: NO  
FRAGMENT TYPE: COMPLETE PEPTIDE  
ORIGINAL SOURCE: SYNTHETIC  
IMMEDIATE SOURCE: SYNTHETIC  
PUBLICATION INFORMATION: NOT PREVIOUSLY PUBLISHED  
US-08-231-730A-27

Query Match 67.4%; Score 29; DB 1; Length 27;  
Best Local Similarity 62.5%; Pred. No. 66;  
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 AKKKKKK 8  
DB 20 AKKKKKK 27

## RESULT 11

US-08-427-001C-27  
Sequence 27, Application US/08427001C  
Patent No. 5717064  
GENERAL INFORMATION:  
APPLICANT: JULIAN, GORDON R.  
TITLE OF INVENTION: METHYLATED LYSINE-RICH LYTIC PEPTIDES.  
NUMBER OF SEQUENCES: 38  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: ROTHELL, FIGG, ERNST & KURZ  
STREET: 555 Thirteenth Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.4 MB STORAGE  
OPERATING SYSTEM: DOS  
SOFTWARE: WordPerfect  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/427,001C  
FILING DATE: 24-Apr-95  
CLASSIFICATION: 530  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: U.S. 08/148,889  
FILING DATE: 08-NOV-93  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: WALKER, BARBARA W.  
REGISTRATION NUMBER: 35,400  
REFERENCE/DOCKET NUMBER: 2093-105A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)783-6040  
TELEFAX: (202)783-6031  
INFORMATION FOR SEQ ID NO: 27:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 27  
TYPE: AMINO ACID  
TOPOLOGY: LINEAR  
MOLECULE TYPE: PEPTIDE  
US-08-427-001C-27

Query Match 67.4%; Score 29; DB 1; Length 27;  
Best Local Similarity 62.5%; Pred. No. 66;  
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 AKKKKKK 8  
DB 20 AKKKKKK 27

## RESULT 12

US-08-457-798-27  
Sequence 27, Application US/08457798  
Patent No. 5744445  
GENERAL INFORMATION:  
APPLICANT: JAYNES, JESSE M.  
TITLE OF INVENTION: METHOD OF TREATING PULMONARY DISEASE  
TITLE OF INVENTION: STATES WITH NON-NATURALLY OCCURRING  
TITLE OF INVENTION: AMPHIPATHIC PEPTIDES

NUMBER OF SEQUENCES: 38  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: JESSE M. JAYNES,  
ADDRESSEE: DEMETER BIOTECHNOLOGIES, LTD.  
STREET: 150 PAVETTEVILLE ST. MALL, SUITE 2700  
CITY: RALEIGH  
STATE: NORTH CAROLINA  
COUNTRY: USA  
ZIP: 27601  
COMPUTER READABLE FORM:  
MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.4 MB STORAGE  
OPERATING SYSTEM: MACINTOSH  
SOFTWARE: M.S. WORD 5.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/457,798  
FILING DATE:  
CLASSIFICATION: 514  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US/08/039,620A  
FILING DATE: 19930604  
ATTORNEY/AGENT INFORMATION:  
NAME: HULTQUIST, STEVEN J.  
REGISTRATION NUMBER: 28021  
REFERENCE/DOCKET NUMBER: 4013-103  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (919)990-9531  
TELEFAX: (919)990-9532  
INFORMATION FOR SEQ ID NO: 27:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 27  
TYPE: AMINO ACID  
TOPOLOGY: LINEAR  
MOLECULE TYPE: PEPTIDE  
DESCRIPTION: NO  
HYPOTHETICAL: NO  
FRAGMENT TYPE: COMPLETE PEPTIDE  
ORIGINAL SOURCE: SYNTHETIC  
IMMEDIATE SOURCE: SYNTHETIC  
PUBLICATION INFORMATION: NOT PREVIOUSLY PUBLISHED  
US-08-457-798-27

Query Match 67.4%; Score 29; DB 1; Length 27;  
Best Local Similarity 62.5%; Pred. No. 66;  
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 AKKKKKK 8  
DB 20 AKKKKKK 27

## RESULT 13

US-08-457-171-27  
Sequence 27, Application US/08457171  
Patent No. 5773413  
GENERAL INFORMATION:  
APPLICANT: JAYNES, JESSE M.  
TITLE OF INVENTION: METHOD OF COMBATING MAMMALIAN NEOPLASIA, AND LYTIC PEPTIDES  
NUMBER OF SEQUENCES: 40  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: STEVEN J. HULTQUIST  
ADDRESS: INTELLECTUAL PROPERTY/TECHNOLOGY LAW  
STREET: 200 PARK DRIVE, SUITE 210  
STREET: P.O. BOX 14329  
CITY: RESEARCH TRIANGLE PARK  
STATE: NORTH CAROLINA  
COUNTRY: USA  
ZIP: 27709  
COMPUTER READABLE FORM:  
MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.4 MB STORAGE  
OPERATING SYSTEM: APPLE MACINTOSH

SOFTWARE: M.S. WORD 5.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/457,171  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/225,476A  
FILING DATE: 04-08-94  
APPLICATION NUMBER: 08/039,620  
FILING DATE: 06-04-93  
APPLICATION NUMBER: 08/148,491  
FILING DATE: 11-08-93  
APPLICATION NUMBER: 08/148,889  
FILING DATE: 11-08-93  
ATTORNEY/AGENT INFORMATION:  
NAME: HULTQUIST, STEVEN J.  
REGISTRATION NUMBER: 28021  
REFERENCE/DOCKET NUMBER: 4013-106  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (919)990-9531  
TELEFAX: (919)990-9532  
INFORMATION FOR SEQ ID NO: 27:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 27  
TYPE: AMINO ACID  
TOPOLOGY: LINEAR  
MOLECULE TYPE:  
DESCRIPTION: PEPTIDE  
HYPOTHETICAL: NO  
FRAGMENT TYPE: COMPLETE PEPTIDE  
ORIGINAL SOURCE: SYNTHETIC  
IMMEDIATE SOURCE: SYNTHETIC  
PUBLICATION INFORMATION: NOT PREVIOUSLY PUBLISHED  
US-08-457-171-27

Query Match 67.4%; Score 29; DB 1; Length 27;  
Best Local Similarity 62.5%; Pred. No. 66;  
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 AKKKKK 8  
|||:|:|  
Db 20 AKKKKK 27

RESULT 14  
US-08-505-486-27  
Sequence 27, Application US/08505486  
Patent No. 5955573  
GENERAL INFORMATION:  
APPLICANT: Jesse M. Jaynes  
TITLE OF INVENTION: UBIQUITIN-LYTIC PEPTIDE FUSION GENE  
TITLE OF INVENTION: CONSTRUCTS, PROTEIN PRODUCTS DERIVING THEREFROM, AND  
TITLE OF INVENTION: METHODS OF MAKING AND USING SAME  
NUMBER OF SEQUENCES: 98  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: ROTHWELL, FIGG, ERNST & KURZ  
STREET: 555 Thirteenth Street N.W.  
CITY: Washington  
STATE: D. C.  
COUNTRY: USA  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.4 MB STORAGE  
COMPUTER: IBM COMPATIBLE  
OPERATING SYSTEM: DOS  
SOFTWARE: WordPerfect 5.1+  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/505,486  
FILING DATE: 21-JUL-1995  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: U.S. 08/279,472  
FILING DATE: 22-JUL-1994  
CLASSIFICATION: 536

ATTORNEY/AGENT INFORMATION:  
NAME: WALKER, BARBARA W.  
REGISTRATION NUMBER: 35,400  
REFERENCE/DOCKET NUMBER: 2093-117A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)783-6040  
TELEFAX: (202)783-6031  
INFORMATION FOR SEQ ID NO: 27:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 27  
TYPE: AMINO ACID  
TOPOLOGY: LINEAR  
MOLECULE TYPE:  
DESCRIPTION: PEPTIDE  
HYPOTHETICAL: NO  
FRAGMENT TYPE: COMPLETE PEPTIDE  
ORIGINAL SOURCE: SYNTHETIC  
IMMEDIATE SOURCE: SYNTHETIC  
PUBLICATION INFORMATION: NOT PREVIOUSLY PUBLISHED  
US-08-505-486-27

Query Match 67.4%; Score 29; DB 2; Length 27;  
Best Local Similarity 62.5%; Pred. No. 66;  
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 AKKKKK 8  
|||:|:|  
Db 20 AKKKKK 27

RESULT 15  
US-08-689-489C-27  
Sequence 27, Application US/08689489C  
Patent No. 6001805  
GENERAL INFORMATION:  
APPLICANT: Jesse M. Jaynes, Gordon R. Julian  
TITLE OF INVENTION: Method of Enhancing Wound Healing By  
TITLE OF INVENTION: Stimulating Fibro-blast and Keratinocyte Growth In  
TITLE OF INVENTION: Vivo, Utilizing Amphipathic Peptides  
NUMBER OF SEQUENCES: 46  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Rothwell, Figg, Ernst & Kurz  
STREET: 555 13TH STREET  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/689,489C  
FILING DATE: August 12, 1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/231,730  
FILING DATE: April 20, 1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/225,476  
FILING DATE: April 8, 1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/039,620  
FILING DATE: June 4, 1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/148,889  
FILING DATE: No. 6001805ember 8, 1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/148,491  
FILING DATE: No. 6001805ember, 8, 1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Mark I. Bowditch  
REGISTRATION NUMBER: 40,315

REFERENCE/DOCKET NUMBER: 2093-120  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-783-6040  
TELEFAX: 202-783-6031  
INFORMATION FOR SEQ. ID NO: 27:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 27 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
FRAGMENT TYPE: linear  
US-08-689-489C-27

Query Match 67.4%; Score 29; DB 3; Length 27;  
Best Local Similarity 62.5%; Pred. No. 66;  
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;  
QY 1 AKMKRRK 8  
|||:|:|  
Db 20 AKVKKKK 27

Search completed: January 29, 2003, 14:04:20  
Job time : 13.0769 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Comphen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2003, 14:00:01 ; Search time 6.92308 Seconds  
(without alignments)  
26.232 Million cell updates/sec

Title: US-09-807-949a-137  
Perfect score: 43  
Sequence: 1 AKMKRRKA 9

Scoring table: BIOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 122226 seqs, 20178551 residues

Total number of hits satisfying chosen parameters: 122226

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database:

Published Applications AA\*  
1: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep:\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep:\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep:\*  
6: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/1/pubpaa/PCTUS09\_PUBCOMB.pep:\*  
8: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09\_PUB.pep:\*  
10: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep:\*  
12: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep:\*  
13: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep:\*  
14: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	33	76.7	218	10	US-09-828-644-99
2	32	74.4	921	9	US-10-153-273-8
3	31	72.1	22	10	US-09-949-196-25
4	31	72.1	22	10	US-09-949-196-29
5	31	72.1	280	12	US-10-115-899-5
6	31	72.1	280	12	US-10-115-899-8
7	31	72.1	592	9	US-09-736-457-1809
8	31	72.1	592	9	US-09-902-941-1809
9	31	72.1	592	9	US-09-849-626-1809
10	30	69.8	13	10	US-09-746-170-2
11	30	69.8	35	12	US-10-001-870-145
12	30	69.8	45	10	US-09-864-761-36386
13	30	69.8	46	10	US-09-864-761-36491
14	30	69.8	348	10	US-09-764-864-1143
15	30	69.8	543	9	US-10-033-297-143
16	30	69.8	550	9	US-09-738-626-3983
17	29	67.4	53	10	US-09-864-761-42895
18	29	67.4	67	12	US-10-001-843-156
19	29	67.4	101	10	US-09-864-761-33599

20	29	67.4	132	10	US-09-925-297-779	Sequence 779, App
21	29	67.4	300	10	US-09-764-870-474	Sequence 474, App
22	29	67.4	335	10	US-09-764-870-314	Sequence 314, App
23	29	67.4	611	10	US-09-833-790-233	Sequence 233, App
24	29	67.4	961	10	US-09-801-368-132	Sequence 132, App
25	28	65.1	22	10	US-09-949-196-37	Sequence 37, Appl
26	28	65.1	22	10	US-09-949-196-41	Sequence 41, Appl
27	28	65.1	38	10	US-09-922-261-222	Sequence 222, App
28	28	65.1	41	10	US-09-922-261-220	Sequence 220, App
29	28	65.1	51	10	US-09-922-261-218	Sequence 218, App
30	28	65.1	57	10	US-09-922-261-216	Sequence 216, App
31	28	65.1	59	10	US-09-864-761-34665	Sequence 34665, A
32	28	65.1	62	10	US-09-922-261-214	Sequence 214, App
33	28	65.1	64	10	US-09-922-261-212	Sequence 212, App
34	28	65.1	70	10	US-09-922-261-210	Sequence 210, App
35	28	65.1	76	10	US-09-922-261-208	Sequence 208, App
36	28	65.1	93	10	US-09-922-261-206	Sequence 206, App
37	28	65.1	101	10	US-09-922-261-204	Sequence 204, App
38	28	65.1	160	9	US-09-991-496-135	Sequence 135, App
39	28	65.1	182	10	US-09-870-162A-19	Sequence 19, Appl
40	28	65.1	210	10	US-09-189-833B-10	Sequence 10, Appl
41	28	65.1	210	10	US-09-902-705-10	Sequence 10, Appl
42	28	65.1	215	10	US-09-143-127-7	Sequence 7, Appl
43	28	65.1	222	10	US-09-143-127-1	Sequence 1, Appl
44	28	65.1	225	10	US-09-143-127-8	Sequence 8, Appl
45	28	65.1	228	10	US-09-143-127-2	Sequence 2, Appl

#### ALIGNMENTS

RESULT 1  
US-09-828-644-99  
Sequence 99, Application US/09828644  
Patent No. US20020015998A1  
GENERAL INFORMATION:  
APPLICANT: Vogell, Gabriel  
TITLE OF INVENTION: No. US20020015998A1 G Protein-Coupled Receptors  
FILE REFERENCE: 00196US1  
CURRENT APPLICATION NUMBER: US/09/828,644  
CURRENT FILING DATE: 2001-04-06  
PRIOR APPLICATION NUMBER: 60/195,150  
PRIOR FILING DATE: 2000-04-06  
PRIOR APPLICATION NUMBER: 60/195,099  
PRIOR FILING DATE: 2000-04-06  
PRIOR APPLICATION NUMBER: 60/195,151  
PRIOR FILING DATE: 2000-04-06  
PRIOR APPLICATION NUMBER: 60/195,148  
PRIOR FILING DATE: 2000-04-06  
PRIOR APPLICATION NUMBER: 60/195,093  
PRIOR FILING DATE: 2000-04-06  
PRIOR APPLICATION NUMBER: 60/195,098  
PRIOR FILING DATE: 2000-04-06  
PRIOR APPLICATION NUMBER: 60/230,149  
PRIOR FILING DATE: 2000-09-05  
NUMBER OF SEQ ID NOS: 117  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 99  
LENGTH: 218  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-828-644-99

Query Match 76.7%; Score 33; DB 10; Length 218,  
Best Local Similarity 75.0%; Pred. No. 48;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 KMKRRKA 9  
Db 53 KMKRRKA 60

RESULT 2

US-10-153-273-8  
; Sequence 8, Application US/10153273  
; Patent No. US20020169305A1  
; GENERAL INFORMATION:  
; APPLICANT: Sim, Kim L.  
; Childs, Chetan  
; Miller, Louis H.  
; Peterson, David S.  
; Su, Xin-zhaun  
; Wellems, Thomas E.  
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX  
; AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS  
; NUMBER OF SEQUENCES: 37  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Knobb Martens Olson & Bear  
; STREET: 620 Newport Center Drive 16th Floor  
; CITY: Newport Beach  
; STATE: California  
; COUNTRY: US  
; ZIP: 92660  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/10/153,273  
; FILING DATE: 21-May-2002  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/09/210,288  
; FILING DATE: <Unknown>  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Puller, Michael  
; REGISTRATION NUMBER: 36,516  
; REFERENCE/DOCKET NUMBER: NIH121.1FMDV1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 235-8550  
; TELEFAX: (619) 235-0176  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 921 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHEICAL: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Plasmodium falciparum  
; SEQUENCE DESCRIPTION: SEQ ID NO: 8:  
US-10-153-273-8  
Query Match 74.4%; Score 32; DB 9; Length 921;  
Best Local Similarity 85.7%; Pred. No. 2.6e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 2 KKKKKR 8  
Db 912 KKKKKR 918  
RESULT 3  
US-09-949-196-25  
; Sequence 25, Application US/09949196  
; Patent No. US20020147145A1  
; GENERAL INFORMATION:  
; APPLICANT: Zealand Pharmaceuticals A/S  
; TITLE OF INVENTION: MATERIALS AND METHODS RELATING TO THE DEGRADATION OF Cdc25A IN RE  
; TITLE OF INVENTION: TO DNA DAMAGE  
; FILE REFERENCE: 55888 (45487)  
; CURRENT APPLICATION NUMBER: US/09/949,196  
; CURRENT FILING DATE: 2001-07-09  
; NUMBER OF SEQ ID NOS: 45

; SOFTWARE: Patentin version 3.1  
; SEQ ID NO 25  
; LENGTH: 22  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: synthetic peptide sequence  
; NAME/KEY: BINDING  
; LOCATION: (22)-(22)  
; OTHER INFORMATION: NH2  
US-09-949-196-25  
Query Match 72.1%; Score 31; DB 10; Length 22;  
Best Local Similarity 66.7%; Pred. No. 12;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
QY 1 AKKKRRRA 9  
Db 9 ARRTRRRA 17  
RESULT 4  
US-09-949-196-29  
; Sequence 29, Application US/09949196  
; Patent No. US20020147145A1  
; GENERAL INFORMATION:  
; APPLICANT: Zealand Pharmaceuticals A/S  
; TITLE OF INVENTION: MATERIALS AND METHODS RELATING TO THE DEGRADATION OF Cdc25A IN RES  
; FILE REFERENCE: 55888 (45487)  
; CURRENT APPLICATION NUMBER: US/09/949,196  
; CURRENT FILING DATE: 2001-07-09  
; NUMBER OF SEQ ID NOS: 45  
; SOFTWARE: Patentin version 3.1  
; SEQ ID NO 29  
; LENGTH: 22  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: synthetic peptide sequence  
; NAME/KEY: BINDING  
; LOCATION: (22)-(22)  
; OTHER INFORMATION: NH2  
US-09-949-196-29  
Query Match 72.1%; Score 31; DB 10; Length 22;  
Best Local Similarity 66.7%; Pred. No. 12;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
QY 1 AKKKRRRA 9  
Db 9 ARRTRRRA 17  
RESULT 5  
US-10-115-899-5  
; Sequence 5, Application US/10115899  
; Patent No. US20020151025A1  
; GENERAL INFORMATION:  
; APPLICANT: Otsuka Pharmaceutical Co., Ltd.  
; TITLE OF INVENTION: Human TSC403 gene and human INCL gene  
; FILE REFERENCE: G60193  
; CURRENT APPLICATION NUMBER: US/10/115,899  
; CURRENT FILING DATE: 2002-04-05  
; PRIOR APPLICATION NUMBER: 09/601,478  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: JP H10-134679  
; PRIOR FILING DATE: 1998-04-28  
; PRIOR APPLICATION NUMBER: JP H10-73234  
; PRIOR FILING DATE: 1998-03-05  
; PRIOR APPLICATION NUMBER: JP H10-38133

;; PRIOR FILING DATE: 1998-02-03  
;; NUMBER OF SEQ ID NOS: 11  
;; SOFTWARE: PatentIn Ver. 2.1  
;; SEQ ID NO 5  
;; LENGTH: 280  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; OTHER INFORMATION: human embryonic brain cDNA library  
US-10-115-899-5

Query Match 72.1%; Score 31; DB 12; Length 280;  
Best Local Similarity 77.8%; Pred. No. 1.3e+02;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AKKKRKA 9  
Db 186 AKKKRKA 194

RESULT 6  
US-10-115-899-8  
; Sequence 8, Application US/10115899  
; Patent No. US20020151025A1  
; GENERAL INFORMATION:  
; APPLICANT: Oetuka Pharmaceutical Co., Ltd.  
; TITLE OF INVENTION: Human TSC403 gene and human INGL gene  
; FILE REFERENCE: 060193  
; CURRENT APPLICATION NUMBER: US/10/115,899  
; PRIOR FILING DATE: 2002-04-05  
; PRIOR APPLICATION NUMBER: 09/601,478  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: JP H10-134679  
; PRIOR FILING DATE: 1998-04-28  
; PRIOR APPLICATION NUMBER: JP H10-73234  
; PRIOR FILING DATE: 1998-03-05  
; PRIOR APPLICATION NUMBER: JP H10-38133  
; PRIOR FILING DATE: 1998-02-03  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 8  
; LENGTH: 280  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; OTHER INFORMATION: human embryonic brain cDNA library  
US-10-115-899-8

Query Match 72.1%; Score 31; DB 12; Length 280;  
Best Local Similarity 77.8%; Pred. No. 1.3e+02;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AKKKRKA 9  
Db 186 AKKKRKA 194

RESULT 7  
US-09-736-457-1809  
; Sequence 1809, Application US/09736457  
; Patent No. US20020168637A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Tongtong  
; APPLICANT: Bangur, Chaitanya S.  
; APPLICANT: Lodes, Michael A.  
; APPLICANT: Fanger, Gary  
; APPLICANT: Vedvick, Tom  
; APPLICANT: Carter, Darick  
; APPLICANT: Retter, Marc  
; APPLICANT: Mannion, Jane  
; APPLICANT: Pan, Aljun  
; APPLICANT: Wang, Aljun  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND  
; TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER

;; FILE REFERENCE: 210121.478C15  
;; CURRENT APPLICATION NUMBER: US/09/736,457  
;; CURRENT FILING DATE: 2000-12-13  
;; NUMBER OF SEQ ID NOS: 1864  
;; SOFTWARE: FastSeq for Windows Version 3.0  
;; SEQ ID NO 1809  
;; LENGTH: 592  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
US-09-736-457-1809

Query Match 72.1%; Score 31; DB 9; Length 592;  
Best Local Similarity 85.7%; Pred. No. 2.5e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 3 KMRKKA 9  
Db 582 KMRKKA 588

RESULT 8  
US-09-902-941-1809  
; Sequence 1809, Application US/09902941  
; Patent No. US20020172952A1  
; GENERAL INFORMATION:  
; APPLICANT: Henderson, Robert A.  
; APPLICANT: Wang, Tongtong  
; APPLICANT: Watanabe, Yoshihiro  
; APPLICANT: Johnson, Jeffrey C.  
; APPLICANT: Retter, Marc W.  
; APPLICANT: Marnerakis, Margarita  
; APPLICANT: Carter, Darick  
; APPLICANT: Fanger, Gary R.  
; APPLICANT: Vedvick, Thomas S.  
; APPLICANT: Bangur, Chaitanya S.  
; APPLICANT: McNabb, Andria  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER  
; FILE REFERENCE: 210121.478C17  
; CURRENT APPLICATION NUMBER: US/09/902,941  
; PRIOR FILING DATE: 2001-07-10  
; NUMBER OF SEQ ID NOS: 2002  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1809  
; LENGTH: 592  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-902-941-1809

Query Match 72.1%; Score 31; DB 9; Length 592;  
Best Local Similarity 85.7%; Pred. No. 2.5e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 3 KMRKKA 9  
Db 582 KMRKKA 588

RESULT 9  
US-09-849-626-1809  
; Sequence 1809, Application US/09849626  
; Publication No. US20020197669A1  
; GENERAL INFORMATION:  
; APPLICANT: Bangur, Chaitanya  
; APPLICANT: Fanger, Gary  
; APPLICANT: Wang, Aljun  
; APPLICANT: Wang, Tongtong  
; APPLICANT: Switzer, Anne  
; APPLICANT: McNeill, Patricia  
; APPLICANT: Clapper, Jonathan  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND  
; TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER  
; FILE REFERENCE: 210121.478C16

; CURRENT APPLICATION NUMBER: US/09/849,626  
; CURRENT FILING DATE: 2001-05-03  
; NUMBER OF SEQ ID NOS: 1926  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 1809  
; LENGTH: 592  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-849-626-1809

Query Match 72.1%; Score 31; DB 9; Length 592;  
Best Local Similarity 85.7%; Pred. No. 2.5e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 KMKRKA 9  
|||  
Db 582 KMKRKA 588

RESULT 10  
US-09-746-170-2  
; Sequence 2, Application US/09746170  
; Patent No. US20020127543A1  
; GENERAL INFORMATION:  
; APPLICANT: Karn, Jonathan  
; APPLICANT: Walker, Stephen  
; TITLE OF INVENTION: Methods and Compositions Utilizing Hepatitis C Virus Molecules  
; FILE REFERENCE: 22620/1280  
; CURRENT APPLICATION NUMBER: US/09/746,170  
; CURRENT FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/171,804  
; PRIOR FILING DATE: 1999-12-22  
; NUMBER OF SEQ ID NOS: 48  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 2  
; LENGTH: 13  
; TYPE: PRT  
; ORGANISM: Hepatitis C virus  
US-09-746-170-2

Query Match 69.8%; Score 30; DB 10; Length 13;  
Best Local Similarity 85.7%; Pred. No. 11;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 KMKRKA 8  
|||  
Db 4 KMKRKA 10

RESULT 11  
US-10-001-870-145  
; Sequence 145, Application US/10001870  
; Patent No. US20020150924A1  
; GENERAL INFORMATION:  
; APPLICANT: Salceda, Susana  
; APPLICANT: Macina, Roberto  
; APPLICANT: Recipon, Herre  
; APPLICANT: Sun, Yongming  
; APPLICANT: Liu, Chenghua  
; TITLE OF INVENTION: Compositions and Methods Relating to Prostate Specific Genes and  
; FILE REFERENCE: DEX-0283  
; CURRENT APPLICATION NUMBER: US/10/001,870  
; CURRENT FILING DATE: 2001-11-20  
; PRIOR APPLICATION NUMBER: 60/252,189  
; PRIOR FILING DATE: 2000-11-21  
; NUMBER OF SEQ ID NOS: 217  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 145  
; LENGTH: 35  
; TYPE: PRT  
; ORGANISM: Homo sapien  
US-10-001-870-145

Query Match 69.8%; Score 30; DB 12; Length 35;  
Best Local Similarity 75.0%; Pred. No. 28;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 AKMKRKA 8  
|||  
Db 4 AKMKRKA 11

RESULT 12  
US-09-864-761-36386  
; Sequence 36386, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharon G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; FILE REFERENCE: Aeomica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263,6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 09/774,203  
; PRIOR FILING DATE: 2001-01-29  
; NUMBER OF SEQ ID NOS: 49117  
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 36386  
; LENGTH: 45  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AC007016.5  
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.5  
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 4.2  
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4.8  
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4.5  
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3.6  
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.9

OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3.8  
 OTHER INFORMATION: EXPRESSED IN HB1100, SIGNAL = 3.7  
 OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 3.4  
 OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.5  
 OTHER INFORMATION: EST\_HUMAN HIT: T06680.1, EVALUATE 8.00e-13  
 OTHER INFORMATION: SWISSPROT HIT: O43101, EVALUATE 4.60e+00  
 US-09-864-761-36386

Query Match 69.8%; Score 30; DB 10; Length 45;  
 Best Local Similarity 66.7%; Pred. No. 35;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AKKKKKRKA 9  
 Db 2 AKKKKKRKA 10

RESULT 13  
 US-09-864-761-36491  
 Sequence 36491, Application US/09864761  
 Patent No. US20020048763A1  
 GENERAL INFORMATION:  
 APPLICANT: Penn, Sharon G.  
 APPLICANT: Rank, David R.  
 APPLICANT: Hanzel, David K.  
 APPLICANT: Chen, Wensheng  
 TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
 TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY  
 FILE REFERENCE: Aemica-X-1  
 CURRENT APPLICATION NUMBER: US/09/864,761  
 CURRENT FILING DATE: 2001-05-23  
 PRIOR APPLICATION NUMBER: US 60/180,312  
 PRIOR FILING DATE: 2000-02-04  
 PRIOR APPLICATION NUMBER: US 60/207,456  
 PRIOR FILING DATE: 2000-05-26  
 PRIOR APPLICATION NUMBER: US 09/632,366  
 PRIOR FILING DATE: 2000-08-03  
 PRIOR APPLICATION NUMBER: GB 24263.6  
 PRIOR FILING DATE: 2000-10-04  
 PRIOR APPLICATION NUMBER: US 60/236,359  
 PRIOR FILING DATE: 2000-09-27  
 PRIOR APPLICATION NUMBER: PCT/US01/00666  
 PRIOR FILING DATE: 2001-01-30  
 PRIOR APPLICATION NUMBER: PCT/US01/00667  
 PRIOR FILING DATE: 2001-01-30  
 PRIOR APPLICATION NUMBER: PCT/US01/00664  
 PRIOR FILING DATE: 2001-01-30  
 PRIOR APPLICATION NUMBER: PCT/US01/00669  
 PRIOR FILING DATE: 2001-01-30  
 PRIOR APPLICATION NUMBER: PCT/US01/00665  
 PRIOR FILING DATE: 2001-01-30  
 PRIOR APPLICATION NUMBER: PCT/US01/00668  
 PRIOR FILING DATE: 2001-01-30  
 PRIOR APPLICATION NUMBER: PCT/US01/00663  
 PRIOR FILING DATE: 2001-01-30  
 PRIOR APPLICATION NUMBER: PCT/US01/00662  
 PRIOR FILING DATE: 2001-01-30  
 PRIOR APPLICATION NUMBER: PCT/US01/00661  
 PRIOR FILING DATE: 2001-01-30  
 PRIOR APPLICATION NUMBER: PCT/US01/00670  
 PRIOR FILING DATE: 2001-01-30  
 PRIOR APPLICATION NUMBER: US 60/234,687  
 PRIOR FILING DATE: 2000-09-21  
 PRIOR APPLICATION NUMBER: US 09/608,408  
 PRIOR FILING DATE: 2000-06-30  
 PRIOR APPLICATION NUMBER: US 09/774,203  
 PRIOR FILING DATE: 2001-01-29  
 NUMBER OF SEQ ID NOS: 49117  
 SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
 SEQ ID NO 36491  
 LENGTH: 46  
 TYPE: PRT  
 ORGANISM: Homo sapiens

FEATURE:  
 OTHER INFORMATION: MAP TO AC012443.1  
 OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.4  
 OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.2  
 OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.7  
 OTHER INFORMATION: EXPRESSED IN HB1100, SIGNAL = 2.5  
 OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.5  
 OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.8  
 OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.7  
 OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.2  
 OTHER INFORMATION: EST\_HUMAN HIT: AUI39074.1, EVALUATE 2.00e-05  
 US-09-864-761-36491

Query Match 69.8%; Score 30; DB 10; Length 46;  
 Best Local Similarity 55.6%; Pred. No. 35;  
 Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AKKKKKRKA 9  
 Db 17 AKKKKKRKA 25

RESULT 14  
 US-09-764-864-1143  
 Sequence 1143, Application US/09764864  
 Patent No. US20020132753A1  
 GENERAL INFORMATION:  
 APPLICANT: Rosen et al.  
 TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
 FILE REFERENCE: PT223  
 CURRENT APPLICATION NUMBER: US/09/764,864  
 CURRENT FILING DATE: 2001-01-17  
 Prior application data removed - consult PAM or file wrapper  
 NUMBER OF SEQ ID NOS: 1792  
 SOFTWARE: Patent In Ver. 2.0  
 SEQ ID NO 1143  
 LENGTH: 348  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 US-09-764-864-1143

Query Match 69.8%; Score 30; DB 10; Length 348;  
 Best Local Similarity 55.6%; Pred. No. 2.2e+02;  
 Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AKKKKKRKA 9  
 Db 185 ARRRRRRRA 193

RESULT 15  
 US-10-033-297-143  
 Sequence 143, Application US/10033297  
 Publication No. US20020187486A1  
 GENERAL INFORMATION:  
 APPLICANT: Hall, Jeff G.  
 APPLICANT: Lyamichev, Victor I.  
 APPLICANT: Mast, Andrea L.  
 APPLICANT: Brow, Mary Ann D.  
 TITLE OF INVENTION: Detection Of Nucleic Acids By Multiple  
 CORRESPONDENCE ADDRESSES:  
 ADDRESSER: Medlen & Carroll, LLP  
 STREET: 220 Montgomery Street, Suite 2200  
 CITY: San Francisco  
 STATE: California  
 COUNTRY: United States Of America  
 ZIP: 94104  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk

```

? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patent Release #1.0, Version #1.30
?
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/10/033,297
? FILING DATE: 12-NO. US20020187486A1-2001
? CLASSIFICATION: <Unknown>
?
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US/09/350,597
? FILING DATE: 09-Jul-1999
? APPLICATION NUMBER: US/08/823,516
? FILING DATE: 24-MAR-1997
? APPLICATION NUMBER: PCT/US97/01072
? FILING DATE: 21-JAN-1997
? APPLICATION NUMBER: US 08/759,038
? FILING DATE: 02-DEC-1996
? APPLICATION NUMBER: US 08/758,314
? FILING DATE: 02-DEC-1996
? APPLICATION NUMBER: US 08/756,386
? FILING DATE: 29-NOV-1996
? APPLICATION NUMBER: US 08/682,853
? FILING DATE: 12-JUL-1996
? APPLICATION NUMBER: US 08/599,491
? FILING DATE: 24-JAN-1996
?
? ATTORNEY/AGENT INFORMATION:
? NAME: Ingolia, Diane E.
? REGISTRATION NUMBER: 40,027
? REFERENCE/DOCKET NUMBER: FORS-02736
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (415) 705-8410
? TELEFAX: (415) 397-8338
?
? INFORMATION FOR SEQ ID NO: 143:
?
? SEQUENCE CHARACTERISTICS:
? LENGTH: 543 amino acids
? TYPE: amino acid
? STRANDEDNESS: NO. US20020187486A1 Relevant
? TOPOLOGY: NO. US20020187486A1 Relevant
? MOLECULE TYPE: protein
? SEQUENCE DESCRIPTION: SEQ ID NO: 143:
?
US-10-033-297-143
?
Query Match 69.8%; Score 30; DB 9; Length 543;
Best Local Similarity 85.7%; Pred. No. 3.3e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 2 KMKRRK 8
DB 534 KMKRRK 540

```

Search completed: January 29, 2003, 14:04:55  
Job time : 6.92308 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd

OM protein - protein search, using sw model

Run on: January 29, 2003, 13:51:05 ; Search time 32.5385 Seconds  
(without alignments)  
36.857 Million cell updates/sec

Title: US-09-807-949A-138

Perfect score: 45  
Sequence: 1 AITFNAQYA 9

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

```
Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries
```

Database :

A\_Geneseq.101002.\*

1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*

2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*

3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*

4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*

5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*

6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*

7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*

8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*

9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*

10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*

11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*

12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*

13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*

14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*

15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*

16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*

17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*

18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*

19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*

20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*

21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*

22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	45	100.0	9	21	AA030356	MN protein CA domain
2	33	73.3	460	23	AB027338	Streptococcus poly
3	33	73.3	462	23	AB027339	Streptococcus poly
4	32	71.1	50	18	AA028348	Staphylococcus aure
5	32	71.1	483	23	AB054838	Lactococcus lactis
6	32	71.1	858	22	AB030216	Novel human diagen
7	32	71.1	867	23	AB035600	Fungal ZBC protein
8	32	71.1	914	22	AB026069	Novel human diagen
9	32	71.1	1014	22	AA034478	Porphyromonas ging
10	32	71.1	1017	20	AA034477	Porphyromonas ging

11	32	71.1	1046	20	AAV34353	Porphyromonas gingiv
12	31	68.9	181	23	AAU6953	Aspergillus niger
13	31	68.9	181	23	AAU6954	Aspergillus niger
14	31	68.9	181	23	AAU6955	Aspergillus niger
15	31	68.9	181	23	AAU6956	Aspergillus niger
16	31	68.9	181	23	AAU6959	Aspergillus niger
17	31	68.9	184	19	AAW60276	Modified xylanase
18	31	68.9	184	19	AAW60728	Xylanase A of Aspe
19	31	68.9	184	21	AAW84554	Aspergillus niger
20	31	68.9	184	21	AAW98050	A. niger xylanase,
21	31	68.9	184	23	AAE18437	Aspergillus niger
22	31	68.9	185	19	AAW60729	Xylanase A of Aspe
23	31	68.9	185	19	AAW60277	Modified xylanase
24	31	68.9	185	19	AAW37825	Amino acid sequenc
25	31	68.9	185	21	AAW98051	A. tubigenensis xyla
26	31	68.9	185	23	AAE18438	Aspergillus tubige
27	31	68.9	211	13	AAW20241	Aspergillus tubige
28	31	68.9	211	13	AAW20241	Aspergillus tubige
29	31	68.9	211	13	AAW20241	Aspergillus prepro
30	31	68.9	211	14	AAW37877	Sequence of prepro
31	31	68.9	211	20	AAW95553	Aspergillus tubige
32	31	68.9	211	21	AAW84552	Aspergillus tubige
33	31	68.9	211	23	AAW69592	Aspergillus niger
34	31	68.9	722	19	AAW55104	Aspergillus niger
35	31	68.9	722	22	ABP54558	Streptococcus pneu
36	31	68.9	1052	22	ABG33543	S. pneumoniae, SP05
37	31	68.9	1052	22	ABG329026	Novel human diagno
38	31	68.9	2234	21	AAW15052	Novel human diagno
39	30	66.7	105	22	AAO00754	Streptococcus pneu
40	30	66.7	211	23	AAU69551	Human polypeptide
41	30	66.7	256	19	AAW98476	Aspergillus niger
42	30	66.7	317	21	AAW84554	H. pylori GHPc 889
43	30	66.7	317	21	AAW39958	Arabidopsis thalia
44	30	66.7	318	21	AAW36233	Arabidopsis thalia
45	30	66.7	318	21	AAW39957	Arabidopsis thalia
	30	66.7	332	17	AAW5449	Arabidopsis thalia
	30	66.7	332	17	AAW5449	Phase abortive inf

## ALIGNMENTS

```

RESULT 1
AAB03056
ID AAB03056 standard; peptide; 9 AA.

```

DT	25-SEP-2000	(first entry)
XX		
DE	MN protein CA domain-blinding peptide, SEQ ID NO:138	

KM MN protein; tumour associated cell adhesion molecule; oncoprotein,  
KM proteoglycan domain; PG domain; carbonic anhydrase; CA domain;  
KM abnormal expression; neoplastic disease; cancer; gene therapy;  
KM phage display library.

OS Synthetic

FH	Key	Location/Qualifiers

FT	Modified-site	9	/note= "C-terminal amide moiety"
FT			

PN WO200024913-A2

PD 04-MAY-2000.

PF 22-OCT-1999; 99WO-US24879.

PR 23-OCT-1998; 98US-0177776.

2000

PA (VIRO-) INST VIROLOGY.

PI	Zavada J	Pastorekova S,	Pastorek J;
DR	WPI, 2000-350752/30.		
PT	A molecule which specifically binds to a site on MN protein (oncoprotein) and prevents adhesion of vertebrate cells to the protein, useful for treating preneoplastic or neoplastic diseases such as cancer		
PT	_		
XX	Clatm 5; Page 68; 154pp; English.		
XX	The invention relates to the inhibition of cell adhesion mediated by the MN oncoprotein (also known as the MN/CA IX isoenzyme or the MN/G250 protein). The MN protein is a tumour-associated adhesion molecule which comprises a proteoglycan-like (PG) domain (AAB03017) which contains the protein's binding site, and a carbonic anhydrase (CA) domain (AAB03018). Abnormal expression of the MN protein is associated with tumorigenicity. The invention encompasses molecules (e.g., proteins and peptides) which which specifically bind to a site on the MN protein, thereby preventing adhesion of vertebrate cells to the protein in a cell adhesion assay. It also encompasses MN proteins or MN protein fragments which can be added to the extracellular environment to prevent the adhesion of vertebrate cells to each other. The invention also relates to the identification of the binding site of the MN protein and to a method of identifying a site on an MN protein to which cells adhere, comprising testing a series of overlapping peptides from the protein in a cell adhesion assay. The invention encompasses a vector comprising an expression control sequence operatively linked to a nucleic acid encoding the variable domains of a MN-specific antibody, where the domains are separated by a flexible linker peptide (AAB03035) and the vector inhibits the growth of a vertebrate preneoplastic or neoplastic cell that abnormally expresses MN protein. The invention also encompasses a vector comprising a nucleic acid encoding a cytotoxic protein or peptide operatively linked to the MN gene promoter, which inhibits the growth of a vertebrate preneoplastic or neoplastic cell. Also claimed is a repressor complex that binds to the MN gene promoter (AAB52473). MN proteins and peptides, MN-binding proteins and peptides, and expression vectors encoding such proteins and peptides are useful for treating patients with preneoplastic or neoplastic disease (e.g., cancers) associated with or characterised by abnormal MN expression. Sequences AAB03032-B03034 and AAB03055-B03058 represent synthetic phage display library peptides which bind to the CA domain of the human MN protein (AAB03005).		
XX			
SO	Sequence	9 AA;	
	Query Match	100.0%; Score 45; DB 21; Length 9;	
	Best Local Similarity	100.0%; Pred. No. 7.8e+05;	
	Matches	9; Conservative 0; Mismatches 0; Indels 0; Gaps 0	
OY	1 AITFNAOYA 9		
Db	1 AITFNAOYA 9		
RESULT 2			
ID	ABP27238	standard; Protein; 460 AA.	
XX	ABP27238;		
AC	ABP27238;		
XX	02-JUL-2002	(first entry)	
DT			
XX	Streptococcus polypeptide SEQ ID NO 3652.		
DE			
XX	Streptococcus; GAS; group B streptococcus; Streptococcus agalactiae;		
KW	group A streptococcus; Streptococcus pyogenes; antibacterial.		
XX	ant inflammatory; infection; vaccine; meningitis; gene therapy.		
XX	Streptococcus agalactiae.		
OS			
XX	MO200234771-AA.		
PN			
XX	02-MAY-2002.		
PD			

PF	29-OCT-2001; 2001WO-GB04789.
XX	
PR	27-OCT-2000; 2000GB-0026333.
XX	
PR	24-NOV-2000; 2000GB-0028727.
XX	
PR	07-MAR-2001; 2001GB-0005640.
XX	
PA	(CHIR-) CHIRON SPA.
PA	(GENO-) INST GENOMIC RES.
PI	Telford J,
PI	Tetzelin H;
XX	
DR	WPI: 2002-352536/38.
XX	N-PSDB; ABN67869.
XX	
PT	New Streptococcus protein for the treatment or prevention of infection
PT	or disease caused by Streptococcus bacteria, such as meningitis, and
PT	for detecting a compound that binds to the protein -
XX	
PS	Claim 1; Page 3520; 4525pp; English.
XX	
CC	The invention relates to a protein (ABP25413-ABP30895) from group B
CC	streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
CC	(Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
CC	the specification. The proteins have antibacterial and anti-inflammatory
CC	activity. (I), nucleic acids encoding (I), ABN65044-ABN71526 and
CC	antibodies that bind (I) are used in the manufacture of medicaments for
CC	the treatment or prevention of infection or disease caused by
CC	Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
CC	Nucleic acids encoding (I) are used to detect Streptococcus in a
CC	biological sample. (I) is used to determine whether a compound binds to
CC	(I). A composition comprising (I) or a nucleic acid encoding (I), may be
CC	used as a vaccine or diagnostic composition. The disease caused by
CC	Streptococcus that is prevented or treated may be meningitis. Nucleic
CC	acid encoding (I) may be used to recombinantly produce (I) and may be
CC	used in gene therapy. Antibodies to (I) are used for affinity
CC	chromatography, immunoassays, and distinguishing/identifying
CC	Streptococcus proteins.
SQ	Sequence 460 AA;
Query Match	73.3%; Score 33; DB 23; Length 460;
Best Local Similarity	62.5%; Pred. No. 1.7e+02;
Matches 5; Conservative	2; Mismatches 1; Indels 0; Gaps 0
OY	1 AITFNAOY R   :   :   62 AVTYNGQY 69
Db	62 AVTYNGQY 69
RESULT 3	
ABP27239	
ID	ABP27239 standard; Protein; 462 AA.
AC	ABP27239;
DT	02-JUL-2002 (first entry)
DE	Streptococcus polypeptide SEQ ID NO 3654.
XX	
KM	Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
KM	group A streptococcus; Streptococcus pyogenes; antibacterial;
KM	antiinflammatory; infection; vaccine; meningitis; gene therapy.
OS	Streptococcus pyogenes.
PN	WO200234771-A2.
PD	02-MAY-2002.
PF	29-OCT-2001; 2001WO-GB04789.
XX	



PR 27-OCT-2000; 2000GB-0026333.  
 PR 24-NOV-2000; 2000GB-0028727.  
 PR 07-MAR-2001; 2001GB-0005640.  
 XX  
 PA (CHIR-) CHIRON SPA.  
 PA (GENO-) INST GENOMIC RES.  
 XX Telford J, Masigiani V, Margarit Ros YI, Grandi G, Fraser C,  
 PI Tettelein H;  
 XX  
 DR WPI, 2002-352536/38.  
 DR N-PSDB; ABN67870.  
 XX  
 PT New Streptococcus protein for the treatment or prevention of infection  
 PT or disease caused by Streptococcus bacteria, such as meningitis, and  
 PT for detecting a compound that binds to the protein -  
 PS Claim 1; Page 3520; 4525pp; English.  
 XX  
 CC The invention relates to a protein (ABP25413-ABP30895) from group B  
 CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS  
 CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in  
 CC the specification. The proteins have antibacterial and anti-inflammatory  
 CC activity. (1), nucleic acids encoding (1), ABN6044-ABN71526 and  
 CC antibodies that bind (1) are used in the manufacture of medicaments for  
 CC the treatment or prevention of infection or disease caused by  
 CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.  
 CC Nucleic acids encoding (1) are used to detect Streptococcus in a  
 CC biological sample. (1) is used to determine whether a compound binds to  
 CC (1). A composition comprising (1) or a nucleic acid encoding (1), may be  
 CC used as a vaccine or diagnostic composition. The disease caused by  
 CC Streptococcus that is prevented or treated may be meningitis. Nucleic  
 CC acid encoding (1) may be used to recombinantly produce (1) and may be  
 CC used in gene therapy. Antibodies to (1) are used for affinity  
 CC chromatography, immunoassays, and distinguishing/identifying  
 CC Streptococcus proteins.  
 CC  
 SQ Sequence 462 AA;  
 Query Match 73.3%; Score 33; DB 23; Length 462;  
 Best Local Similarity 62.5%; Pred. No. 1.7e+02;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 OY 1 ATFFNAGY 8  
 Db 62 AVTYNGCY 69  
 RESULT 4  
 AAM28348  
 ID AAM28348 standard; Protein; 50 AA.  
 AC AAM28348;  
 XX  
 DT 01-SBP-1998 (first entry)  
 XX  
 DE Staphylococcus aureus protein of unknown function.  
 XX  
 KW Staphylococcus aureus protein; ribozyme; antisense sequence; control;  
 KW Staphylococcal gene; regulatory element; bacterial gene expression;  
 KW vaccine; Staphylococcal infection; food poisoning; scaled skin syndrome;  
 KW toxic shock syndrome.  
 XX  
 OS Staphylococcus aureus.  
 XX  
 PH Key Location/Qualifiers  
 FT Misc-difference 3  
 FT Misc-difference 3 /note= "X is not defined in the specification"  
 FT Misc-difference 4 /note= "X is not defined in the specification"  
 PT  
 XX  
 PN W09730070-A1.  
 XX

PD 21-AUG-1997.  
 XX  
 PF 19-FEB-1997; 97WO-US02318.  
 XX  
 XX  
 PR 20-FEB-1996; 96US-0011888.  
 XX  
 PA (SMIK ) SMITHKLINE BEECHAM CORP.  
 XX  
 PI Black MT, Burnham MK, Hodgson JE, Knowles DJC, Nicholas RO;  
 PI Pratt JM, Reichard RW, Rosenberg M, Ward JM;  
 XX  
 DR WPI, 1997-424969/39.  
 DR N-PSDB; AAT84238.  
 XX  
 PT Novel polypeptide(s) from Staphylococcus aureus strain WCUH29 - used  
 PT to isolate antimicrobial compounds, and in vaccines against S.  
 PT aureus infection  
 PS Claim 6; Page 624; 989pp; English.  
 XX  
 CC The present sequence represents a Staphylococcus aureus protein of  
 CC unknown function. The DNA sequence was isolated from a library of  
 CC clones of S. aureus WCUH 29 in Escherichia coli. The DNA sequence can  
 CC be used in the construction of ribozymes and antisense sequences to  
 CC control the expression of Staphylococcal genes. The DNA sequence is  
 CC also useful as a source of regulatory elements for the control of  
 CC bacterial gene expression. The present protein may be used to produce  
 CC vaccines to enable a host to produce specific antibodies with  
 CC antibacterial action. These vaccines and antibodies would protect  
 CC a host against invasion by S. aureus, and conditions relating to  
 CC Staphylococcal infection, e.g. Staphylococcal food poisoning, scaled  
 CC skin syndrome, and toxic shock syndrome.  
 CC  
 SQ Sequence 50 AA;  
 Query Match 71.1%; Score 32; DB 18; Length 50;  
 Best Local Similarity 77.8%; Pred. No. 27;  
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 OY 1 ATFFNAGYA 9  
 Db 16 ALPFLALYA 24  
 RESULT 5  
 ABB54838  
 ID ABB54838 standard; Protein; 483 AA.  
 AC ABB54838;  
 XX  
 DT 16-MAY-2002 (first entry)  
 XX  
 DE Lactococcus lactis protein ypcG.  
 XX  
 KW Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese.  
 XX  
 OS Lactococcus lactis IL1403.  
 XX  
 PN FR2807446-A1.  
 XX  
 PD 12-OCT-2001.  
 XX  
 PF 11-APR-2000; 2000FR-0004630.  
 XX  
 PR 11-APR-2000; 2000FR-0004630.  
 XX  
 PA (INRG ) INRA INST NAT RECH AGRONOMIQUE.  
 XX  
 PI Bolotine A, Sorokine A, Renault P, Ehrlich SD;  
 XX  
 DR WPI, 2002-043418/06.  
 XX  
 PT New nucleotide sequence useful in the identification of Lactococcus

PT Lactis and related species -  
 XX  
 PS Claim 6; SEQ ID No 1540; 2504bp; French.  
 XX  
 CC The present invention is related to a *Lactococcus lactis* nucleotide  
 CC sequence (AB950521) and related proteins (AB95300-AB955621). The  
 CC nucleic acid sequence is useful in the detection and/or amplification of  
 CC nucleic acid sequence, particularly to identify *Lactococcus lactis* or  
 CC related species. The proteins of the invention are useful for the  
 CC biosynthesis or biodegradation of a composition of interest. The  
 CC invention helps research in lactic bacteria, particularly useful in the  
 CC production of yogurt and cheese.  
 CC Note: The sequence data for this patent is based on equivalent patent  
 CC WO20017734 (published 18-OCT-2001) which is available in electronic  
 CC format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 483 AA;  
 Query Match 71.1%; Score 32; DB 23; Length 483;  
 Best Local Similarity 85.7%; Pred. No. 2.9e+02;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 2 ITFNAQY 8  
 Db 157 ITFNAQY 163  
 RESULT 6  
 ABG30216  
 ID ABG30216 standard; Protein; 858 AA.  
 XX  
 AC ABG30216;  
 XX  
 DT 18-FEB-2002 (first entry)  
 XX  
 DE Novel human diagnostic protein #30207.  
 XX  
 KM Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KM food supplement; medical imaging; diagnostic; genetic disorder.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200175067-A2.  
 XX  
 PD 11-OCT-2001.  
 XX  
 PF 30-MAR-2001; 2001WO-US08631.  
 XX  
 PR 31-MAR-2000; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Dzmanac RT, Liu C, Tang YT;  
 XX  
 DR WPI; 2001-639362/73.  
 DR N-PSDB; AAS94403.  
 XX  
 PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -  
 XX  
 PS Claim 20; SEQ ID No 60575; 103bp; English.  
 XX  
 CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridization probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or

CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 858 AA;  
 Query Match 71.1%; Score 32; DB 22; Length 858;  
 Best Local Similarity 85.7%; Pred. No. 5.2e+02;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 3 TENAQYA 9  
 Db 450 TENASTA 456  
 RESULT 7  
 ABP35600  
 ID ABP35600 standard; Protein; 867 AA.  
 XX  
 AC ABP35600;  
 XX  
 DT 24-JUL-2002 (first entry)  
 XX  
 DE Fungal ZBC protein sequence #26.  
 XX  
 KM Secondary metabolite; fungus; ZBC gene; zinc binuclear cluster protein;  
 KM antibacterial; beta-lactam; anti-hypercholesterolemic; lovastatin;  
 KM mevastatin; immunosuppressant; cyclosporin A; ergot alkaloid; ergotamine;  
 KM angiogenesis inhibitor; ovalicin; glucan synthase inhibitor; gliotoxin;  
 KM fungal toxin; cell surface receptor; plant growth regulator; pigment;  
 KM insecticide; antineoplastic.  
 XX  
 OS Unidentified.  
 XX  
 PN WO200224865-A2.  
 XX  
 PD 28-MAR-2002.  
 XX  
 PF 19-SEP-2001; 2001WO-US29288.  
 XX  
 PR 19-SEP-2000; 2000US-233564P.  
 XX  
 PA (MICR-) MICROBIA INC.  
 XX  
 PI Holtzman D, Madden K, Maxon M, Sherman A;  
 XX  
 DR WPI; 2002-352005/38.  
 DR N-PSDB; ABN79789.  
 XX  
 PT New method for improving the production of a secondary metabolite e.g.  
 PT antineoplastic agent, ergot alkaloid from a fungus involves modulation  
 PT of the expression of at least one zinc binuclear cluster protein gene  
 XX  
 PS Disclosure; SEQ ID 64; 49bp + sequence listing; English.  
 XX  
 CC The invention relates to improving the production of a secondary  
 CC metabolite by a fungus. This involves modulating the expression of at  
 CC least one ZBC (zinc binuclear cluster protein) gene in a manner to  
 CC improve the yield of the secondary metabolite. Methods of the invention  
 CC may be used for improving the production of the secondary metabolite e.g.  
 CC antibacterial (such as beta-lactam), an anti-hypercholesterolemic (such  
 CC as lovastatin or mevastatin), an immunosuppressant (such as cyclosporin A),

CC an ergot alkaloid (such as ergotamine), an angiogenesis inhibitor (such  
CC as ovalicin), a glucan synthase inhibitor, gliotoxin family of compounds,  
CC a fungal toxin, a modulator of cell surface receptor signalling, a plant  
CC growth regulator, a pigment, an insecticide, or an antineoplastic  
CC compound. The method results in a decrease in fermentor run-time, a  
CC decrease in the size of the fermentor required for the production of  
CC equivalent amounts of the secondary metabolite, or a decrease in the  
CC biomass required for the production, which translates into decreased  
CC waste that must be handled in downstream processing. The sequences given  
CC in records AB35575-AB35722 represent ZBC protein. The sequences given  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.

CC Sequence 867 AA;

Query Match 71.1%; Score 32; DB 23; Length 867;  
Best Local Similarity 62.5%; Pred. No. 5.2e+02;  
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 2 ITFNAQYA 9  
Db 206 LYVNAQFA 213

RESULT 8

ABG26069  
ID ABG26069 standard; Protein; 914 AA.

XX ABG26069;

DT 18-FEB-2002 (first entry)

DE Novel human diagnostic protein #26060.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
XX food supplement; medical imaging; diagnostic; genetic disorder.

OS Homo sapiens.

XX WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

XX 23-AUG-2000; 2000US-0649167.

PA (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;

DR WPI; 2001-639362/73.

DR N-PSDB; AAS90256.

PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -

PS Claim 20; SEQ ID No 56428; 103bp; English.

XX The invention relates to isolated polynucleotide (I) and  
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,  
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
XX and gene mapping, and in recombinant production of (II). The  
XX polynucleotides are also used in diagnostics as expressed sequence tags  
XX for identifying expressed genes. (I) is useful in gene therapy techniques  
XX to restore normal activity of (II) or to treat disease states involving  
XX (II). (II) is useful for generating antibodies against it, detecting or  
XX quantitating a polypeptide in tissue, as molecular weight markers and as  
XX a food supplement. (II) and its binding partners are useful in medical

CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG0010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

CC Sequence 914 AA;

Query Match 71.1%; Score 32; DB 22; Length 914;  
Best Local Similarity 85.7%; Pred. No. 5.5e+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 TFNAQYA 9  
Db 506 TFNASYA 512

RESULT 9

AAV34478  
ID AAV34478 standard; Protein; 1014 AA.

XX AAV34478;

DT 25-AUG-1999 (first entry)

DE Porphyromonas gingivalis protein PG2.

XX Porphyromonas gingivalis; PG; periodontal disease; gingivitis;  
XX vaccine; antigenic.

OS Porphyromonas gingivalis.

XX WO9929870-A1.

PD 17-JUN-1999.

PF 10-DEC-1998; 98WO-AU01023.

PR 04-AUG-1998; 98AU-0005028.

PR 10-DEC-1997; 97AU-0000839.

PR 31-DEC-1997; 97AU-0001182.

PR 30-JAN-1998; 98AU-0001546.

PR 10-MAR-1998; 98AU-0002264.

PR 09-APR-1998; 98AU-0002911.

PR 23-APR-1998; 98AU-0003128.

PR 05-MAY-1998; 98AU-0003338.

PR 22-MAY-1998; 98AU-0003654.

PR 29-JUL-1998; 98AU-0004917.

PA (CSLC-) CSL LTD.

PI Agius CT, Barr TG, Hocking DM, Margetts MB, Patterson MA;

PI Ross BC, Rothel LJ, Webb EA;

DR WPI; 1999-385613/32.

DR N-PSDB; AAX91696.

PT Antigenic Porphyromonas gingivalis peptides for preventing  
PT gingivitis  
XX Claim 1, Page 458-460; 588bp; English.

XX AAX91536 to AAX91801 encode two hundred and sixty six antigenic  
XX Porphyromonas gingivalis (Pg) polypeptide sequences given in AAX94318 to  
XX AAX94583. AAX91802 to AAX91989 represent PCR primers used in the  
XX isolation of the Pg polypeptides. The Pg polypeptides have antibacterial  
XX activity with a vaccine mechanism of action. The Pg polypeptides can be

CC used as vaccines especially against Porphyromonas gingivalis. Probes can  
CC be used to detect Porphyromonas gingivalis in standard hybridisation  
CC assays. Porphyromonas gingivalis is involved in periodontal disease  
CC especially gingivitis.

XX Sequence 1014 AA;

SO Query Match 71.1%; Score 32; DB 20; Length 1014;  
Best Local Similarity 85.7%; Pred. No. 6.2e+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 ITFNAY 8  
Db 236 ITFNASY 242

RESULT 10  
AAV34477  
ID AAV34477 standard; Protein; 1017 AA.

AC AAV34477;

DT 25-AUG-1999 (first entry)

DE Porphyromonas gingivalis protein PG2.

XX Porphyromonas gingivalis; PG; periodontal disease; gingivitis;  
KM vaccine; antigenic.

XX Porphyromonas gingivalis.

PN WO929870-A1.

PD 17-JUN-1999.

PF 10-DEC-1998; 98WO-AU01023.

PR 04-AUG-1998; 98AU-0005028.

PR 10-DEC-1997; 97AU-0000839.

PR 31-DEC-1997; 97AU-0001182.

PR 30-JAN-1998; 98AU-0001546.

PR 10-MAR-1998; 98AU-0002264.

PR 09-APR-1998; 98AU-0002911.

PR 23-APR-1998; 98AU-0003128.

PR 05-MAY-1998; 98AU-0003338.

PR 22-MAY-1998; 98AU-0003654.

PR 29-JUL-1998; 98AU-0004917.

PA (CSLC-) CSL LTD.

PI Agius CT, Barr IG, Hocking DM, Margetts MB, Patterson MA;  
PI Ross BC, Rothel LJ, Webb EA;

DR WPI; 1999-385613/32.

XX N-PSDB; AAX91695.

PT Antigenic Porphyromonas gingivalis peptides for preventing  
PT gingivitis

PS Claim 1; Page 456-458; 588pp; English.

XX AAX91536 to AAX91801 encode two hundred and sixty six antigenic

CC Porphyromonas gingivalis (PG) polypeptide sequences given in AAV34318 to

CC AAX34583. AAX91802 to AAX91989 represent PCR primers used in the

CC isolation of the PG polypeptides. The PG polypeptides have antibacterial

CC activity with a vaccine mechanism of action. The PG polypeptides can be

CC used as vaccines especially against Porphyromonas gingivalis. Probes can

CC be used to detect Porphyromonas gingivalis in standard hybridisation

CC assays. Porphyromonas gingivalis is involved in periodontal disease

CC especially gingivitis.

Query Match 71.1%; Score 32; DB 20; Length 1017;  
Best Local Similarity 85.7%; Pred. No. 6.2e+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 ITFNAY 8  
Db 239 ITFNASY 245

RESULT 11  
AAV34353  
ID AAV34353 standard; Protein; 1046 AA.

AC AAV34353;

DT 25-AUG-1999 (first entry)

DE Porphyromonas gingivalis protein PG2.

XX Porphyromonas gingivalis; PG; periodontal disease; gingivitis;  
KM vaccine; antigenic.

XX Porphyromonas gingivalis.

PN WO929870-A1.

PD 17-JUN-1999.

PF 10-DEC-1998; 98WO-AU01023.

PR 04-AUG-1998; 98AU-0005028.

PR 10-DEC-1997; 97AU-0000839.

PR 31-DEC-1997; 97AU-0001182.

PR 30-JAN-1998; 98AU-0001546.

PR 10-MAR-1998; 98AU-0002264.

PR 09-APR-1998; 98AU-0002911.

PR 23-APR-1998; 98AU-0003128.

PR 05-MAY-1998; 98AU-0003338.

PR 22-MAY-1998; 98AU-0003654.

PR 29-JUL-1998; 98AU-0004917.

PA (CSLC-) CSL LTD.

PI Agius CT, Barr IG, Hocking DM, Margetts MB, Patterson MA;  
PI Ross BC, Rothel LJ, Webb EA;

DR WPI; 1999-385613/32.

XX N-PSDB; AAX91571.

PT Antigenic Porphyromonas gingivalis peptides for preventing  
PT gingivitis

PS Claim 1; Page 314-316; 588pp; English.

XX AAX91536 to AAX91801 encode two hundred and sixty six antigenic

CC Porphyromonas gingivalis (PG) polypeptide sequences given in AAV34318 to

CC AAX34583. AAX91802 to AAX91989 represent PCR primers used in the

CC isolation of the PG polypeptides. The PG polypeptides have antibacterial

CC activity with a vaccine mechanism of action. The PG polypeptides can be

CC used as vaccines especially against Porphyromonas gingivalis. Probes can

CC be used to detect Porphyromonas gingivalis in standard hybridisation

CC assays. Porphyromonas gingivalis is involved in periodontal disease

CC especially gingivitis.

SO Sequence 1046 AA;

Query Match 71.1%; Score 32; DB 20; Length 1046;

Best Local Similarity 85.7%; Pred. No. 6.4e+02;

Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 ITFNAY 8  
Db 268 ITFNASY 274

```

RESULT 12
ID AAU96953 standard; Protein; 181 AA.
XX
XX AAU96953;
AC
DT 30-JUL-2002 (first entry)
XX
DE Aspergillus niger xylanase wild-type protein sequence.
XX
XX Plant antifungal; fungal xylanase; thermostability; xylan; cellulose;
KM alcoholic liquid; beer; wine; fruit juice; vegetable juice; agriculture;
KW recycling; paper; foodstuff; coffee; plant oil; starch; plant pulp.
XX
OS Aspergillus niger.
XX
XX EPI184460-AL.
FN
PD 06-MAR-2002.
XX
XX 29-AUG-2000; 2000EP-0307374.
PP
XX 29-AUG-2000; 2000EP-0307374.
PR
XX (STAM ) DSM NV.
PA
PI Van Den Hombergh JPTW, Van Der Laan JM, Menke HH, Daran JG;
XX WPI; 2002-332040/37.
DR
XX Protein for, e.g. treating plant material, comprises fungal xylanase
PT modified at exposed serine residues or within positions 90-160
XX
XX Example 2; Fig 6; 74pp; English.
PS
XX The present invention relates to a new protein comprising a fungal
CC xylanase modified to increase thermostability. The modification is at
CC exposed serine residues or within positions 90-160. The invention is
CC used in degrading xylan in cellulose, treating plant material, improving
CC filterability and/or reducing viscosity of xylan-containing liquids,
CC improving filterability or clarifying alcoholic liquids (e.g. beer, wine)
CC or fruit or vegetable juices, hydrolysing agricultural residues, in
CC recycling materials (e.g. containing paper) in papermaking, for
CC thickening foodstuffs, and/or extracting desirable materials (e.g.
CC coffee, plant oil, starch), processing plant pulp, juice or extract.
CC The invention provides a novel protein comprising modified xylanase that
CC is up to 10 times or more stable than its unmodified counterparts, and
CC yet retains a pH optimum that is acidic. The present amino acid sequence
CC represents the Aspergillus niger xylanase wild-type protein of the
CC invention, as described above.
XX
SQ Sequence 181 AA;

```

XX	Plant antifungal; fungal xylanase; thermostability; xylan; cellulose;
XX	alcoholic liquid; beer; wine; fruit juice; vegetable juice; agriculture;
KW	recycling; paper; foodstuff; coffee; plant oil; starch; plant pulp;
KW	mutant; mutein.
XX	
OS	<i>Aspergillus niger</i> .
XX	Synthetic.
XX	
Key	Location/Qualifiers
FT	Misc-difference 104
FT	/note= "Wild-type Ser substituted by Cys"
FT	Misc-difference 148
FT	/note= "Wild-type Asn substituted by Cys"
XX	
XX	EP184460-A1.
XX	
PD	06-MAR-2002.
XX	
PF	29-AUG-2000; 2000EP-0307374.
XX	
PR	29-AUG-2000; 2000EP-0307374.
XX	
PA	(STAM ) DSM NV.
XX	
P1	Van Den Hombergh JPTW, Van Der Laan JM, Menke HH, Daran JG;
DR	WPI; 2002-332040/37.
XX	
PT	Protein for, e.g. treating plant material, comprises fungal xylanase
PT	modified at exposed serine residues or within positions 90-160
XX	
PS	Example 2; Page -: 74pp; English.
XX	
CC	The present invention relates to a new protein comprising a fungal
CC	xylanase modified to increase thermostability. The modification is at
CC	exposed serine residues or within positions 90-160. The invention is at
CC	used in degrading xylan in cellulose, treating plant material, improving
CC	filterability and/or reducing viscosity of xylan-containing liquids,
CC	improving filterability or clarifying alcoholic liquids (e.g. beer, wine)
CC	or fruit or vegetable juices, hydrolysing agricultural residues, in
CC	recycling materials (e.g. containing paper) in papermaking, for
CC	thickening foodstuffs, and/or extracting desirable materials (e.g.
CC	coffee, plant oil, starch), processing plant pulp, juice or extract.
CC	The invention provides a novel protein comprising modified xylanase that
CC	is up to 10 times or more stable than its unmodified counterparts, and
CC	yet retains a pH optimum that is acidic. The present amino acid sequence
CC	represents the <i>Aspergillus niger</i> xylanase St03C/N147C mutant protein
CC	sequence of the invention.
CC	Note: This sequence is not shown in the specification but is derived
CC	from the wild-type <i>Aspergillus niger</i> xylanase protein sequence (AAU96953)
CC	shown in figure 6 of the specification.
XX	
SQ	Sequence 181 AA;
XX	
Query Match	68.9%; Score 31; DB 23; Length 181;
Best Local Similarity	55.6%; Pred. No. 1.6e+02;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;	
QY	1 AITFNAQYA 9
	:
DB	49 AITVSAEYS 57
XX	
RESULT 14	
AAU96955	
ID	AAU96955 standard; Protein, 181 AA.
XX	
AC	AAU96955;
XX	
DT	30-JUL-2002 (first entry)
XX	
DE	<i>Aspergillus niger</i> xylanase A146S mutant protein sequence.

```

XX  plant antifungal; fungal xylanase; thermostability; xylan; cellulose;
KM  alcoholic liquid; beer; wine; fruit juice; vegetable juice; agriculture;
KM  recycling; paper; foodstuff; coffee; plant oil; starch; plant pulp;
KM  mutant; muteln.
OS  Aspergillus niger.
XX  Synthetic.
XX  Key
FH  Location/Qualifiers
FT  Misc-difference 144
   /note= "Wild-type Ala substituted by Ser"
XX  EPI184460-A1.
XX  06-MAR-2002.
XX  29-AUG-2000; 2000EP-0307374.
XX  29-AUG-2000; 2000EP-0307374.
XX  (STAM ) DSM NV.
XX  (STAM ) DSM NV.
PI  Van Den Hombergh JPTW, Van Der Laan JM, Menke HH, Daran JG;
DR  WPI; 2002-332040/37.
XX  Protein for, e.g. treating plant material, comprises fungal xylanase
PT  modified at exposed serine residues or within positions 90-160 -
XX  Example 2; Page -: 74pp; English.
XX  The present invention relates to a new protein comprising a fungal
CC  xylanase modified to increase thermostability. The modification is at
CC  exposed serine residues or within positions 90-160. The invention is
CC  used in degrading xylan in cellulose, treating plant material, improving
CC  filterability and/or reducing viscosity of xylan-containing liquids,
CC  improving filterability or clarifying alcoholic liquids (e.g. beer, wine)
CC  or fruit or vegetable juices, hydrolysing agricultural residues, in
CC  recycling materials (e.g. containing paper) in papermaking, for
CC  thickening foodstuffs, and/or extracting desirable materials (e.g.
CC  coffee, plant oil, starch), processing plant pulp, juice or extract.
CC  The invention provides a novel protein comprising modified xylanase that
CC  is up to 10 times or more stable than its unmodified counterparts, and
CC  yet retains a pH optimum that is acidic. The present amino acid sequence
CC  represents the Aspergillus niger xylanase A146N mutant protein sequence
CC  of the invention.
CC  Note: This sequence is not shown in the specification but is derived
CC  from the wild-type Aspergillus niger xylanase protein sequence (AAU96953)
CC  shown in figure 6 of the specification.
XX  Sequence 181 AA;
SQ
Query Match 68.9%; Score 31; DB 23; Length 181;
Best Local Similarity 55.6%; Pred. No. 1.6e+02;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
OY 1 AITPNAOYA 9
   |||::||:
DB 49 AITSAEYS 57

```

```

KM  alcoholic liquid; beer; wine; fruit juice; vegetable juice; agriculture;
KM  recycling; paper; foodstuff; coffee; plant oil; starch; plant pulp;
KM  mutant; muteln.
OS  Aspergillus niger.
XX  Synthetic.
XX  Key
FH  Location/Qualifiers
FT  Misc-difference 144
   /note= "Wild-type Ala substituted by Asn"
XX  EPI184460-A1.
XX  06-MAR-2002.
XX  29-AUG-2000; 2000EP-0307374.
XX  29-AUG-2000; 2000EP-0307374.
XX  (STAM ) DSM NV.
XX  (STAM ) DSM NV.
PI  Van Den Hombergh JPTW, Van Der Laan JM, Menke HH, Daran JG;
DR  WPI; 2002-332040/37.
XX  Protein for, e.g. treating plant material, comprises fungal xylanase
PT  modified at exposed serine residues or within positions 90-160 -
XX  Example 2; Page -: 74pp; English.
XX  The present invention relates to a new protein comprising a fungal
CC  xylanase modified to increase thermostability. The modification is at
CC  exposed serine residues or within positions 90-160. The invention is
CC  used in degrading xylan in cellulose, treating plant material, improving
CC  filterability and/or reducing viscosity of xylan-containing liquids,
CC  improving filterability or clarifying alcoholic liquids (e.g. beer, wine)
CC  or fruit or vegetable juices, hydrolysing agricultural residues, in
CC  recycling materials (e.g. containing paper) in papermaking, for
CC  thickening foodstuffs, and/or extracting desirable materials (e.g.
CC  coffee, plant oil, starch), processing plant pulp, juice or extract.
CC  The invention provides a novel protein comprising modified xylanase that
CC  is up to 10 times or more stable than its unmodified counterparts, and
CC  yet retains a pH optimum that is acidic. The present amino acid sequence
CC  represents the Aspergillus niger xylanase A146N mutant protein sequence
CC  of the invention.
CC  Note: This sequence is not shown in the specification but is derived
CC  from the wild-type Aspergillus niger xylanase protein sequence (AAU96953)
CC  shown in figure 6 of the specification.
XX  Sequence 181 AA;
SQ
Query Match 68.9%; Score 31; DB 23; Length 181;
Best Local Similarity 55.6%; Pred. No. 1.6e+02;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
OY 1 AITPNAOYA 9
   |||::||:
DB 49 AITSAEYS 57

```

Search completed: January 29, 2003, 13:59:55  
 Job time : 34.5385 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2003, 13:57:01 ; Search time 12.6923 Seconds  
(without alignments)  
68.168 Million cell updates/sec

Title: US-09-807-949a-138  
Perfect score: 45  
Sequence: 1 AITPNAQYA 9

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : PIR 73:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	34	75.6	255	2 T36778	probable enoyl-(ac
2	33	73.3	462	2 G95233	Cof Family protein
3	33	73.3	472	2 A99098	conserved hypochet
4	32	71.1	250	2 AEO151	conserved hypochet
5	32	71.1	307	2 E95939	probable inosine-u
6	32	71.1	356	1 A97033	probable phosphos
7	32	71.1	483	2 G86811	sugar ABC transpor
8	32	71.1	520	2 F89786	conserved hypochet
9	32	71.1	592	2 E89772	hypothetical prote
10	32	71.1	767	2 B95301	conserved hypochet
11	32	71.1	867	2 T41308	hypothetical zinc-
12	31	68.9	211	1 S48229	endo-1,4-beta-xyla
13	31	68.9	211	2 S48542	endo-1,4-beta-xyla
14	31	68.9	260	2 T22676	hypothetical prote
15	31	68.9	335	2 G90276	catechol 2,3-dioxy
16	31	68.9	383	2 JC2472	brain and reproduc
17	31	68.9	421	2 D69643	imidazole-5-prop
18	31	68.9	2228	2 E97942	beta-galactosidase
19	31	68.9	2233	2 B95075	beta-galactosidase
20	30	66.7	152	2 A69806	conserved hypochet
21	30	66.7	171	2 T49493	hypothetical prote
22	30	66.7	175	2 B82716	ATP synthase, delt
23	30	66.7	225	2 B65127	type 4 prepilin-1
24	30	66.7	225	2 H64686	hypothetical prote
25	30	66.7	253	2 A71829	hypothetical prote
26	30	66.7	303	2 B96909	probable permease
27	30	66.7	312	2 D83689	hypothetical prote
28	30	66.7	329	2 T04244	hypothetical prote
29	30	66.7	370	2 C40585	recf protein - Sta

30	30	66.7	403	2 F90283	hypothetical prote
31	30	66.7	570	1 Q0B32R	BRAF1 (BC-RF2) pro
32	30	66.7	572	2 D97342	dihydroxyacid dehyd
33	30	66.7	750	2 T04010	hypothetical prote
34	30	66.7	1306	2 T28313	ORF MSV152 probabl
35	30	66.7	1322	2 T15689	hypothetical prote
36	30	66.7	1335	2 T17508	glycoprotein Vp260
37	30	66.7	4639	1 A54794	dynein heavy chain
38	29	64.4	298	2 T34259	hypothetical prote
39	29	64.4	321	2 H81373	probable integral
40	29	64.4	325	2 A47003	cytokine receptor
41	29	64.4	349	2 JC6311	interferon recepto
42	29	64.4	379	2 G97772	DNA-directed DNA p
43	29	64.4	380	2 B86625	DNA-directed DNA p
44	29	64.4	381	2 B85060	hypothetical prote
45	29	64.4	393	2 A11311	aspartate aminotra

#### ALIGNMENTS

##### RESULT 1

T36778

probable enoyl-(acyl-carrier-protein) reductase - Streptomyces coelicolor

C:Species: Streptomyces coelicolor

C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 31-Jan-2000

C:Accession: T36778

R:Seeger, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.

submitted to the EMBL Data Library, July 1999

A:Reference number: 221574

A:Accession: T36778

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-255 <SEB>

A:Cross-references: EMBL:AL096844, PDB: CAB50882.1, GSPDB: GN00070, SCODEB: SCT128.08c

A:Experimental source: strain A3(2)

A:Genetics:

A:Gene: inhA; SCODEB: SCT128.08c

C:Superfamily: enoyl-(acyl-carrier-protein) reductase (NADH); short-chain alcohol dehydro

Query Match 75.6%; Score 34; DB 2; Length 255;  
Best Local Similarity 75.0%; Pred. No. 9.9;

Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 ITPNAQYA 9

DB 143 LTFDAQYA 150

##### RESULT 2

G95233

Cof Family protein [imported] - Streptococcus pneumoniae (strain TIGR4)

C:Species: Streptococcus pneumoniae

C>Date: 03-Aug-2001 #sequence\_revision 03-Aug-2001 #text\_change 03-Aug-2001

C:Accession: G95233

R:Retselein, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heidt

on, J.D.; Unayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapfle, I

neon, T.; Hickey, E.K.; Holt, I.E.

Science 293, 498-506, 2001

A:Authors: Loftus, B.D.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,

A>Title: Complete genome sequence of a virulent isolate of Streptococcus pneumoniae.

A:Reference number: A95000; MUID:21357209; PMID:11463916

A:Accession: G95233

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-462 <KUR>

A:Cross-references: GB:AB005672; PDB: 1AK76064.1; PID: G14973505; GSPDB: GN00164; TIGR: SP44

A:Experimental source: strain TIGR4

C:Genetics:

A:Gene: SP1997

Query Match 73.3%; Score 33; DB 2; Length 462;

Best Local Similarity 62.5%; Pred. No. 31;

Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 AITFNAOY 8  
 Db 62 AVTNGOT 69

## RESULT 3

A99098

conserved hypothetical protein spr1811 [imported] - Streptococcus pneumoniae (strain R6)  
 C/Species: Streptococcus pneumoniae  
 C/Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 22-Oct-2001

A/Accession: A99098  
 R/Hoskins, J.A.; Albom Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; E  
 r, R.; Leblanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McWhren, S.; M  
 y, P.; Sun, P.M.; Winkler, M.E.  
 J. Bacteriol. 183, 5709-5717, 2001  
 A/Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;  
 A/Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.  
 A/Reference number: A97872; MUID:21429245; PMID:11544234

A/Accession: A99098  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-472 <KUR>  
 A/Cross-references: GB:AE007317; PIDN:PAL00614.1; PID:G15459497; GSPDB:GN00174  
 C/Genetics:  
 A/Gene: spr1811

Query Match 73.3%; Score 33; DB 2; Length 472;  
 Best Local Similarity 62.5%; Pred. No. 32;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 AITFNAOY 8  
 Db 72 AVTNGOT 79

## RESULT 4

AE0151

conserved hypothetical protein YP01238 [imported] - Yersinia pestis (strain CO92)

C/Species: Yersinia pestis  
 C/Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 09-Nov-2001  
 C/Accession: AE0151  
 R/Parikh, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.;  
 deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Dougan, G.;  
 ll, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrett,  
 Nature 413, 523-527, 2001  
 A/Title: Genome sequence of Yersinia pestis, the causative agent of plague.  
 A/Reference number: AB0001; MUID:21470413; PMID:11586360

A/Accession: AE0151  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-250 <KUR>  
 A/Cross-references: GB:AL590842; PIDN:CAC90072.1; PID:G15979292; GSPDB:GN00175  
 C/Genetics:  
 A/Gene: YP01238

C/Superfamily: conserved hypothetical protein H1013

Query Match 71.1%; Score 32; DB 2; Length 250;  
 Best Local Similarity 66.7%; Pred. No. 26;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 AITFNAOY 9  
 Db 98 SITFDEOYA 106

## RESULT 5

E95939

probable inosine-uridine preferring nucleoside hydrolase protein [imported] - Sinorhizob  
 C/Species: Sinorhizobium meliloti  
 C/Date: 24-Aug-2001 #sequence\_revision 24-Aug-2001 #text\_change 02-Nov-2001  
 C/Accession: E95939

R/Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Hernan  
 Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001  
 A/Title: The complete sequence of the 1.683-kb pSymB megaplasmid from the N2-fixing endo  
 A/Reference number: A95842; MUID:21356508; PMID:11481431

A/Accession: E95939  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-307 <KUR>  
 A/Cross-references: GB:AL591985; PIDN:CAC49181.1; PID:G15140666; GSPDB:GN00167  
 A/Experimental source: strain 1021, megaplasmid pSymB  
 R/Gallier, F.; Finan, T.M.; Long, S.R.; Punter, A.; Abola, P.; Ampe, F.; Barloy-Hubler,  
 pel, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federle, N.A.; Fisher, R.F.;  
 L.; Hyman, R.W.; Jones, T.  
 Science 293, 668-672, 2001

A/Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kise, E.; Komp, C.; LeLaure,  
 heault, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Walla, D.H.; Wong, K.; Yeh, K.  
 A/Title: The composite genome of the legume symbiont Sinorhizobium meliloti.  
 A/Reference number: A96039; MUID:21368234; PMID:11474104  
 A/Contents: annotation  
 C/Genetics:  
 A/Gene: SMD21277  
 A/Genome: plasmid  
 C/Superfamily: yafF protein

Query Match 71.1%; Score 32; DB 2; Length 307;  
 Best Local Similarity 85.7%; Pred. No. 33;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 TFINAOYA 9  
 Db 279 TFINAHYA 285

## RESULT 6

A97033

probable phosphoesterase (EC 3.1.-.-) CAC1078 [similarity] - Clostridium acetobutylicum

C/Species: Clostridium acetobutylicum  
 C/Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 22-Oct-2001  
 C/Accession: A97033  
 R/Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,  
 J. Daily, M.J.; Benneit, G.N.; Koonin, E.V.; Smith, D.R.  
 J. Bacteriol. 183, 4823-4838, 2001

A/Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo  
 A/Reference number: A96900; MUID:21359325; PMID:21359325

A/Accession: A97033  
 A/Molecule type: DNA  
 A/Residues: 1-356 <KUR>  
 A/Cross-references: GB:AE001437; PIDN:AAK79052.1; PID:G15023992; GSPDB:GN00168  
 A/Experimental source: Clostridium acetobutylicum ATCC824  
 C/Comment: This sequence has motifs characteristic of a variety of phosphoesterases.  
 C/Genetics:  
 A/Gene: CAC1078  
 C/Superfamily: Clostridium acetobutylicum probable phosphoesterase CAC1078; phosphoester  
 C/Keywords: hydrolase  
 F/33-127/Domain: phosphoesterase core homology <PBC>

Query Match 71.1%; Score 32; DB 1; Length 356;  
 Best Local Similarity 66.7%; Pred. No. 38;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 AITFNAOYA 9  
 Db 25 SITFNAQYS 33

## RESULT 7

G86811

sugar ABC transporter substrate binding protein yocG [imported] - Lactococcus lactis subsp

C/Species: Lactococcus lactis subsp. lactis  
 C/Date: 23-Mar-2001 #sequence\_revision 23-Mar-2001 #text\_change 03-Aug-2001  
 C/Accession: G86811  
 R/Bolotin, A.; Winkler, P.; Mauger, S.; Jallion, O.; Malarme, K.; Weissenbach, J.; Ehrlich  
 Genome Res. 11, 731-753, 2001



A>Title: The complete genome sequence of the lactic acid bacterium *Lactococcus lactis* s8  
A:Reference number: A86625; MUID:21235186; PMID:11337471  
A:Accession: G86811  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-483 <STG>  
A:Cross-references: GB:AE005176; PID:g12724491; PIDN:AAK05593.1; GSPDB:GN00146  
A:Experimental source: strain IL1403  
C:Genetics:  
A:Gene: ypcG

Query Match 71.1%; Score 32; DB 2; Length 483;  
Best Local Similarity 85.7%; Pred. No. 53;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 ITFNAQYA 8  
DB 157 ITFNAQY 163

RESULT 8  
F89786  
conserved hypothetical protein SA0227 [imported] - *Staphylococcus aureus* (strain N315)  
C:Species: *Staphylococcus aureus*  
C>Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 22-Oct-2001  
A:Accession: F89786  
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc  
ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;  
C.; Shiba, T.; Hatori, M.; Ogasawara, N.; Hayashi, H.; Hiramoto, K.  
Lancet 357, 1225-1240, 2001  
A>Title: Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*.  
A:Reference number: A89758; MUID:21311952; PMID:11418146  
A:Accession: F89786  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-520 <KUR>  
A:Cross-references: GB:BA000018; PID:g13700150; PIDN:BA841449.1; GSPDB:GN00149  
A:Experimental source: strain N315  
C:Genetics:  
A:Gene: SA0227

Query Match 71.1%; Score 32; DB 2; Length 520;  
Best Local Similarity 75.0%; Pred. No. 58;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 ITFNAQYA 9  
DB 439 IDFNAQYS 446

RESULT 9  
E89772  
hypothetical protein SA0117 [imported] - *Staphylococcus aureus* (strain N315)  
C:Species: *Staphylococcus aureus*  
C>Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 03-Jun-2002  
A:Accession: E89772  
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc  
ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;  
C.; Shiba, T.; Hatori, M.; Ogasawara, N.; Hayashi, H.; Hiramoto, K.  
Lancet 357, 1225-1240, 2001  
A>Title: Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*.  
A:Reference number: A89758; MUID:21311952; PMID:11418146  
A:Accession: E89772  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-592 <KUR>  
A:Cross-references: GB:BA000018; PID:g13700037; PIDN:BA841336.1; GSPDB:GN00149  
A:Experimental source: strain N315  
C:Genetics:  
A:Gene: SA0117  
C:Superfamily: *Escherichia coli* plasmid ColV-K30 aerobactin biosynthesis protein *incC*

Query Match 71.1%; Score 32; DB 2; Length 592;

Best Local Similarity 77.8%; Pred. No. 67;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATFNAQYA 9  
DB 371 ATFNAQYA 379

RESULT 10  
B95301  
conserved hypothetical protein SMA0601 [imported] - *Sinorhizobium meliloti* (strain 1021)  
C:Species: *Sinorhizobium meliloti*  
C>Date: 24-Aug-2001 #sequence\_revision 24-Aug-2001 #text\_change 30-Sep-2001  
A:Accession: B95301  
R:Barnett, M.J.; Fisher, R.F.; Jones, T.; Komp, C.; Abola, A.P.; Barlow-Hubler, F.; Bowne  
., Kaiman, S.; Keating, D.H.; Palm, C.; Peck, M.C.; Surzycki, R.; Wells, D.H.; Yeh, K.C.;  
Proc. Natl. Acad. Sci. U.S.A. 98, 9883-9888, 2001  
A>Title: Nucleotide sequence and predicted functions of the entire *Sinorhizobium meliloti*  
A:Reference number: A95262; MUID:21396509; PMID:11481432  
A:Accession: B95301  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-767 <KUR>  
A:Cross-references: GB:AE006469; PIDN:AAK4972.1; PID:g14523397; GSPDB:GN00165  
A:Experimental source: strain 1021, megaplasmid pSymA  
R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barlow-Hubler,  
pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Drano, S.; Federspiel, N.A.; Fisher, R.F.;  
L.; Hyman, R.W.; Jones, T.  
Science 293, 666-672, 2001  
A:Authors: Kahn, D.; Kahn, M.L.; Kaiman, S.; Keating, D.H.; Komp, C.; Lelaure,  
heault, P.; Vandenberg, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.C.  
A>Title: The composite genome of the legume symbiont *Sinorhizobium meliloti*.  
A:Reference number: A96039; MUID:21368234; PMID:11474104  
A:Contents: annotation  
C:Genetics:  
A:Gene: SMA0601  
A:Genome: plasmid

Query Match 71.1%; Score 32; DB 2; Length 767;  
Best Local Similarity 66.7%; Pred. No. 88;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATFNAQYA 9  
DB 591 ATFNAQYA 599

RESULT 11  
T41308  
hypothetical zinc-finger protein - fission yeast (*Schizosaccharomyces pombe*)  
C:Species: *Schizosaccharomyces pombe*  
C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 02-Sep-2000  
A:Accession: T41308  
R:Wood, V.; Rajandream, M.A.; Barrett, B.G.; Wedler, H.; Wambutt, R.; Wedler, E.  
submitted to the EMBL Data Library, March 1998  
A:Reference number: Z21986  
A:Accession: T41308  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-867 <WOO>  
A:Cross-references: EMBL:AL022245; PIDN:CA118305.1; GSPDB:GN00068; SPDB:SPCC320.03  
A:Experimental source: strain 972h-; cosmid c320  
C:Genetics:  
A:Gene: SPDB:SPCC320.03  
A:Map position: 3  
C:Superfamily: GAL4 zinc binuclear cluster homology  
F/71-113/Domain: GAL4 zinc binuclear cluster homology <GL4>

Query Match 71.1%; Score 32; DB 2; Length 867;  
Best Local Similarity 62.5%; Pred. No. 1e+02;  
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 ITFNAQYA 9

Db 206 LTYNAQFA 213

# RESULT 12

S48229  
 endo-1,4-beta-xylanase (EC 3.2.1.8) 1A precursor - *Aspergillus awamori*  
 N:Alternate names: xylanase 1A  
 C:Species: *Aspergillus awamori*  
 C>Date: 15-Jul-1995 #sequence\_revision 22-Nov-1996 #text\_change 18-Jun-1999  
 C:Accession: S48229, S43015  
 R:Heising, J.G.M.; van Rotterdam, C.; Verhake, J.M.A.; Roza, M.; Maat, J.; van Gorcom, C.  
 Curr. Genet. 26: 228-232, 1994  
 A:Title: Isolation and characterization of a 1,4-beta-endoxylanase gene of *A. awamori*.  
 A:Reference number: S48229; MUID:95163119; PMID:7859305  
 A:Accession: S48229  
 A:Molecule type: DNA  
 A:Residues: 1-211 <DEG>  
 A:Cross-references: EMBL:X78115; NID:G460340; PIDN:CAAS5005.1; PID:G460341  
 A:Experimental source: ATCC11358  
 C:Genetics:  
 A:Gene: ex1A  
 A:Introns: 77/2  
 C:Function:  
 A:Description: catalyzes the hydrolysis of 1,4-beta-xylosidic bonds in xylans  
 A:Pathway: xylan degradation  
 C:Superfamily: endo-1,4-beta-xylanase; endo-1,4-beta-xylanase homology  
 C:Keywords: glycosidase; hydrolase; polysaccharide degradation  
 F:1-27/Domain: signal sequence #status predicted <SIG>  
 F:28-211/Product: endo-1,4-beta-xylanase 1A #status predicted <MAT>  
 F:30-210/Domain: endo-1,4-beta-xylanase homology <XLT>  
 F:106,197/Active site: Glu #status predicted

Query March 68.9%; Score 31; DB 1; Length 211;  
 Best Local Similarity 55.6%; Pred. No. 36;  
 Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Oy 1 AITNAOYA 9  
 |||||:  
 Db 78 AITSAEYS 86

# RESULT 13

S49542  
 endo-1,4-beta-xylanase (EC 3.2.1.8) A precursor - *Aspergillus tubigenensis*  
 C:Species: *Aspergillus tubigenensis*  
 C>Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 22-Jun-1999  
 C:Accession: S49542; S77906; S77907  
 R:de Graaff, L.H.; van den Broeck, H.C.; van Ooijen, A.J.J.; Vlasser, J.  
 Mol. Microbiol. 12, 479-490, 1994  
 A:Title: Regulation of the xylanase-encoding xlnA gene of *Aspergillus tubigenensis*.  
 A:Reference number: S49542; MUID:94344036; PMID:8065265  
 A:Accession: S49542  
 A:Molecule type: DNA  
 A:Residues: 1-211 <DEG>  
 A:Cross-references: EMBL:L26986; NID:G508305; PIDN:AA805996.1; PID:G1490411  
 A:Experimental source: strain NW756  
 A:Accession: S77906  
 A:Molecule type: mRNA  
 A:Residues: 1-211 <DEG2>  
 A:Cross-references: EMBL:L26986; NID:G508305; PIDN:AA805996.1; PID:G1490411  
 A:Experimental source: strain NW756  
 A:Accession: S77907  
 A:Molecule type: Protein  
 A:Residues: 28-38;107-120 <DEG3>  
 C:Genetics:  
 A:Gene: xlnA  
 A:Introns: 77/2  
 C:Superfamily: endo-1,4-beta-xylanase; endo-1,4-beta-xylanase homology  
 C:Keywords: glycosidase; hydrolase; polysaccharide degradation  
 F:1-27/Domain: signal sequence #status predicted <SIG>  
 F:28-211/Product: endo-1,4-beta-xylanase A #status experimental <MAT>  
 F:30-210/Domain: endo-1,4-beta-xylanase homology <XLT>

F:97,108,142/Binding site: substrate (Tyr, Tyr, Arg) #status predicted  
 F:106,197/Active site: Glu #status predicted  
 F:146-147/Cleavage site: Pro-Ser (unidentified protease) #status predicted

Query Match 68.9%; Score 31; DB 2; Length 211;  
 Best Local Similarity 55.6%; Pred. No. 36;  
 Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Oy 1 AITNAOYA 9  
 |||||:  
 Db 78 AITSAEYS 86

# RESULT 14

T22676  
 hypothetical protein F54F3.4 - *Caenorhabditis elegans*  
 C:Species: *Caenorhabditis elegans*  
 C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 31-Jan-2000  
 C:Accession: T22676; T24959  
 R:Percy, C.; Lloyd, C.  
 submitted to the EMBL Data Library, September 1996  
 A:Reference number: Z19598  
 A:Accession: T22676  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-260 <NTL>  
 A:Cross-references: EMBL:T27966; PIDN:CAB01974.1; GSPDB:GN00023; CESP:F54F3.4  
 A:Experimental source: clone F54F3  
 R:Lloyd, C.  
 submitted to the EMBL Data Library, November 1996  
 A:Reference number: Z19960  
 A:Accession: T24959  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-260 <NT2>  
 A:Cross-references: EMBL:Z81592; PIDN:CAB04734.1; GSPDB:GN00023; CESP:F54F3.4  
 A:Experimental source: clone T16G1  
 C:Genetics:  
 A:Gene: CESP:F54F3.4  
 A:Map position: 5  
 A:Introns: 21/2; 81/3; 201/3  
 C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology

Query Match 68.9%; Score 31; DB 2; Length 260;  
 Best Local Similarity 66.7%; Pred. No. 45;  
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Oy 1 AITNAOYA 9  
 |||||:  
 Db 142 AITNAEYS 150

# RESULT 15

G90276  
 catechol 2,3-dioxygenase (metapyrocatechase) [imported] - *Sulfolobus solfataricus*  
 C:Species: *Sulfolobus solfataricus*  
 C>Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 15-Jun-2001  
 C:Accession: G90276  
 R:Shen, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.V.; Chan-V  
 Jung, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, P  
 arrett, R.A.; Ragan, M.A.; Jensen, C.W.; Van der Oost, J.  
 submitted to GenBank, April 2001  
 A:Description: *Sulfolobus solfataricus* complete genome.  
 A:Reference number: A99139  
 A:Accession: G90276  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-335 <KUR>  
 A:Cross-references: GB:AE006641; NID:G13814422; PIDN:AAK1470.1; GSPDB:GN00155  
 C:Genetics:  
 A:Gene: SS01223  
 C:Superfamily: catechol 2,3-dioxygenase

Query Match 68.9%; Score 31; DB 2; Length 335;  
Best Local Similarity 62.5%; Pred. No. 59;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
QY 2 ITFNAQYA 9  
:|||||  
Db 220 VTFNVDYA 227

Search completed: January 29, 2003, 14:03:27  
Job time : 13.6923 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: January 29, 2003, 13:51:40 ; Search time 6.23077 Seconds  
(without alignments)  
59.910 Million cell updates/sec

Title: US-09-807-949A-138  
Perfect score: 45  
Sequence: 1 ATENAGYA 9

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	32	71.1	262 1 TRYT_DROER	P54628 drosophila
2	31	68.9	211 1 XYNI_ASPAW	P55328 aspergillus
3	31	68.9	211 1 XYNI_ASPNG	P55329 aspergillus
4	31	68.9	211 1 XYNI_ASPGU	P55331 aspergillus
5	31	68.9	421 1 HUTI_BACSU	P42084 bacillus su
6	30	66.7	225 1 LEP4_ECOLI	P25860 escherichia
7	30	66.7	370 1 RECF_STEAM	P29232 streptococ
8	30	66.7	570 1 UL25_BEV	P03333 equus cabal
9	30	66.7	604 1 PGH2_HORSE	O19183 equus cabal
10	30	66.7	4639 1 YUHC_DROME	P37276 drosophila
11	29	64.4	216 1 YUHT_YERHU	O87570 yerinia ru
12	29	64.4	310 1 HUTI_STRGC	P42558 streptococ
13	29	64.4	321 1 Y987_CAMOB	P45390 campylobact
14	29	64.4	325 1 I108_HUMAN	O08339 homo sapien
15	29	64.4	349 1 I108_MOUSE	O61190 mus musculu
16	29	64.4	379 1 DP3B_RICCN	O92137 rickettsia
17	29	64.4	380 1 DP3B_LACIA	O92131 lactococcus
18	29	64.4	421 1 HUTI_STRPY	P58080 streptococ
19	29	64.4	590 1 INRI_MOUSE	P33896 mus musculu
20	29	64.4	946 1 AMPR_PUXY	P91887 pullulla xy
21	29	64.4	1564 1 PDRA_YEAST	P51533 saccharomyc
22	29	64.4	1861 1 APU_THETU	P38536 t amylolupul
23	28	62.2	232 1 YPFH_ECOLI	P76561 escherichia
24	28	62.2	319 1 PAI_SERLI	P18952 serratia li
25	28	62.2	323 1 TKGA_ERWHE	P58004 erwina her
26	28	62.2	334 1 CATL_RAT	P07154 rattus norv
27	28	62.2	354 1 YR98_CAUCR	O9492 caulobacter
28	28	62.2	363 1 YCDM_BCO57	O8492 escherichia
29	28	62.2	380 1 RFC_SHIDY	O03584 shigella dy
30	28	62.2	404 1 NAME_CLOSO	P16868 clostridium
31	28	62.2	406 1 G59F_DROME	O91965 drosophila
32	28	62.2	416 1 IRI2_HCMVA	P16810 human cytom
33	28	62.2	425 1 Y655_ARCFU	O29602 archaeoglob

34	28	62.2	436 1 MDA1_BACSU	P70965 bacillus su
35	28	62.2	463 1 Y030_NPVAC	P41434 autographa
36	28	62.2	469 1 DPD2_MOUSE	O35654 mus musculu
37	28	62.2	504 1 Y795_METHA	O58205 methanococ
38	28	62.2	524 1 GUA1_CORAM	O52831 corynebacte
39	28	62.2	525 1 GUA1_MYCTU	O50729 mycobacteri
40	28	62.2	526 1 GUA1_STRCO	O910h2 streptomyc
41	28	62.2	529 1 GUA1_MYCLE	P46810 mycobacteri
42	28	62.2	550 1 SYN_CHILM	O9D107 chlamydia t
43	28	62.2	550 1 SYN_CHILM	O84035 chlamydia t
44	28	62.2	562 1 LCB2_KLUIA	P48241 kluyveromyc
45	28	62.2	563 1 P072_HUMAN	P30038 homo sapien

## ALIGNMENTS

```

RESULT 1
TRYT_DROER
ID TRYT_DROER STANDARD; PRT; 262 AA.
AC P54628;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Trypsin theta precursor (EC 3.4.21.4).
GN TRY-THETA.
OS Drosophila erecta (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7220;
RN [1]
RP SEQUENCE FROM N.A.
RA Wang S., Hickey D.A.;
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-Xaa, Lys-|-Xaa.
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; U40653; AAA83238.1; -.
CC HSSP; P00761; 1EPT.
DR MEROPS; S01.112; -.
DR Flybase; FBgn0015082; Dere\Try-theta.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser protease_try.
DR Pfam; PF00089; trypsin_1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; TRYD_SPC; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HTS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydroxylase; Serine protease; Zymogen; Signal; Multigene family.
FT SIGNAL 1 19
FT PROPEP 20 34
FT CHAIN 35 262
FT ACT_SITE 76 76
FT ACT_SITE 121 121
FT ACT_SITE 216 216
FT DISULFID 61 77
FT DISULFID 166 203
FT DISULFID 212 236
FT SITE 210 210
FT SITE 210 210
SQ SEQUENCE 262 AA; 28210 MW; EF3BCAD1143E25D CRC64;
Query Match 71.1%; Score 32; DB 1; Length 262;

```

Best Local Similarity 55.6%; Pred. No. 8.3;  
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 AITRNOYA 9  
DB 106 ALTYNADYS 114

## RESULT 2

XYNI\_ASPAW STANDARD; PRT; 211 AA.  
ID XINI\_ASPAW  
AC P55328; Q12534; (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Endo-1,4-beta-xylanase I precursor (EC 3.2.1.8) (Xylanase I)  
DE (1,4-beta-D-xylan xylanohydrolase I).  
GN XYNA OR EXIA.  
OS Aspergillus awamori.  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
OC Eurotiiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.  
NCBI\_TaxID=105351;

SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.

RC STRAIN=ATCC 11358;  
RX MEDLINE=9516319; PubMed=7859305;  
RA Hestring J.G.M., van Rotterdam C.O., Verhake J.M.A., Roza M., Maat J.,  
RA van Gorcom R.F.M., van den Hondel C.A.M.J.S.,  
RT "Isolation and characterization of a 1,4-beta-endoxylanase gene of A.  
RT awamori".  
RL Curr. Genet. 26:228-233 (1994).

CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-xylosidic  
CC linkages in xylans.

CC -1- PATHWAY: Xylan degradation.

CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY G (FAMILY 11 OF GLYCOSYL  
CC HYDROLASES).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
DR EMBL, X78115; CAAS5005.1; -.  
DR HSSP, P33557; BKU.  
DR InterPro; IPR001137; GH\_11.  
DR Pfam; PF00457; Glyco\_hydro\_11; 1.  
DR PRINTS; PR00911; GLHYDRASE1.  
DR PROSITE; PS00776; GLYCOSYL\_HYDROL\_F11\_1; 1.  
DR PROSITE; PS00777; GLYCOSYL\_HYDROL\_F11\_2; 1.  
DR Xylan degradation; Hydrolase; Glycosidase; Signal.  
KW Xylan degradation; Hydrolase; Glycosidase; Signal.  
FT SIGNAL 1 27  
FT CHAIN 28 211  
FT ACT\_SITE 106 106 ENDO-1,4-BETA-XYLANASE I.  
FT ACT\_SITE 197 197 NUCLEOPHILE (BY SIMILARITY).  
FT ACT\_SITE 197 197 PROTON DONOR (BY SIMILARITY).  
SQ SEQUENCE 211 AA; 22627 MW; 86FBE12A869022 CRC64;

Query Match 68.9%; Score 31; DB 1; Length 211;  
Best Local Similarity 55.6%; Pred. No. 11;  
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 AITRNOYA 9  
DB 78 ALTYSABYS 86

## RESULT 3

XYNI\_ASPNG STANDARD; PRT; 211 AA.  
ID XINI\_ASPNG  
AC P55329;  
DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Endo-1,4-beta-xylanase I precursor (EC 3.2.1.8) (Xylanase I)  
DE (1,4-beta-D-xylan xylanohydrolase I).  
GN XYNA.  
OS Aspergillus niger.  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
OC Eurotiiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.  
NCBI\_TaxID=5061;

RN [1]  
RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS).  
RX MEDLINE=97045991; PubMed=8890913;  
RA Krengel U., Dijkstra B.W.;  
RT "Three-dimensional structure of Endo-1,4-beta-xylanase I from  
RT Aspergillus niger: molecular basis for its low pH optimum.";  
RL J. Mol. Biol. 263:70-78 (1996).

CC -1- FUNCTION: HAS A LOW PH OPTIMUM (3.0).  
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-xylosidic  
CC linkages in xylans.

CC -1- PATHWAY: Xylan degradation.

CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY G (FAMILY 11 OF GLYCOSYL  
CC HYDROLASES).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
DR EMBL, A19535; CAA01470.1; -.  
DR PDB; 1UKR; 24-DEC-97.  
DR InterPro; IPR001137; GH\_11.  
DR Pfam; PF00457; Glyco\_hydro\_11; 1.  
DR PRINTS; PR00911; GLHYDRASE1.  
DR PROSITE; PS00776; GLYCOSYL\_HYDROL\_F11\_1; 1.  
DR PROSITE; PS00777; GLYCOSYL\_HYDROL\_F11\_2; 1.  
KW Xylan degradation; Hydrolase; Glycosidase; Signal; 3D-structure.  
FT SIGNAL 1 27  
FT CHAIN 28 211  
FT ACT\_SITE 106 106 ENDO-1,4-BETA-XYLANASE I.  
FT ACT\_SITE 197 197 NUCLEOPHILE.  
FT ACT\_SITE 197 197 PROTON DONOR.  
FT DISULFID 119 138  
SQ SEQUENCE 211 AA; 22641 MW; 82BBE12ED79303 CRC64;

Query Match 68.9%; Score 31; DB 1; Length 211;  
Best Local Similarity 55.6%; Pred. No. 11;  
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 AITRNOYA 9  
DB 78 ALTYSABYS 86

## RESULT 4

XYNI\_ASPTU STANDARD; PRT; 211 AA.  
ID XINI\_ASPTU  
AC P55331; Q12568;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 15-JUN-1998 (Rel. 36, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Endo-1,4-beta-xylanase I precursor (EC 3.2.1.8) (Xylanase I)  
DE (1,4-beta-D-xylan xylanohydrolase I).  
GN XYNA OR XLNA.  
OS Aspergillus tubingensis.  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
OC Eurotiiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.  
NCBI\_TaxID=5068;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MW756;  
RX MEDLINE=94344036; PubMed=8065265;

RA de Graaf L.H., van den Broeck H.C., van Ooijen A.J.J., Visser J.;  
 RT "Regulation of the xylanase-encoding xlnA gene of *Aspergillus*  
 RL *tubigenis*."; Mol. Microbiol. 12:479-490(1994).  
 RP REVISIONS.  
 RA de Graaf L.H., van den Broeck H.C., van Ooijen A.J.J., Visser J.;  
 RL Submitted (Aug-1996) to the EMBL/GenBank/DBJ databases.  
 CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-xylopyranosidic  
 CC linkages in xylans.  
 CC -1- PATHWAY: Xylan degradation.  
 CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY G (FAMILY 11 OF GLYCOSYL  
 CC HYDROLASES).  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; L26988; AAB05996.1; -;  
 DR HSRP; P55329; 10KR.  
 DR InterPro; IPR001137; GH 11.  
 DR Pfam; PF00457; GLYCO\_hydro\_11; 1.  
 DR PRINTS; PR00911; GHYDRPLASE11.  
 DR PROSITE; PS00776; GLYCOSTL\_HYDROL\_F11\_1; 1.  
 DR PROSITE; PS00777; GLYCOSTL\_HYDROL\_F11\_2; 1.  
 DR Xylan degradation; Hydrolase; Glycosidase; Signal.  
 FT SIGNAL 1 27  
 FT CHAIN 28 211  
 FT ACT\_SITE 106 106 ENDO-1,4-BETA-XYLANASE 1.  
 FT ACT\_SITE 197 197 NUCLEOPHILE (BY SIMILARITY).  
 FT ACT\_SITE 197 197 PROTON DONOR (BY SIMILARITY).  
 SQ SEQUENCE 211 AA; 22576 MW; 1A8BD06C57080D4 CRC64;  
 Query Match 68.9%; Score 31; DB 1; Length 211;  
 Best Local Similarity 55.6%; Pred. No. 11;  
 Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 AITFNAOYA 9  
 Db 78 AITFSAEYS 86  
 RESULT 5  
 ID HUT1\_BACSU STANDARD; PRT; 421 AA.  
 AC P42084;  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Imidazole-5-prolone-5-hydroxylase (EC 3.5.2.7) (Imidazole-5-prolone-5-hydroxylase).  
 GN HUT1 OR EES7B.  
 OS *Bacillus subtilis*.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OK NCBI\_TaxId=1423;  
 RN 11  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168 / BGSCJAL;  
 RX MEDLINE=95219088; PubMed=7704263;  
 RA Yoshida K.-I., Sano H., Seki S., Oda M., Fujimura M., Fujita Y.;  
 RT "Cloning and sequencing of a 29 kb region of the *Bacillus subtilis*  
 RL genome containing the hut and wpa loci."; Microbiology 141:337-343(1995).  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE=98044033; PubMed=9384377;  
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,  
 RA Azevedo V., Bertolo M.G., Bessieres P., Bolotin A., Borchert S.,  
 RA Borstis R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,

RA Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,  
 RA Choi S.K., Codani J.J., Connerion I.F., Cummings N.J., Daniel R.A.,  
 RA Denzot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,  
 RA Entian K.D., Errington J., Fabet C., Ferrari E., Foulger D.,  
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,  
 RA Ghm S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,  
 RA Giuseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Henauf A.,  
 RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Iraya M., Jones L.,  
 RA Joris B., Karmata D., Kasahara Y., Klaer-blanchard M., Klein C.,  
 RA Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,  
 RA Kurita K., Lapidus A., Lardinois S., Lauder J., Lazarevic V.,  
 RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,  
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,  
 RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudeg B., Park S.H.,  
 RA Parro V., Pohl T.M., Portelle D., Porwollik S., Prescott A.M.,  
 RA Priesean E., Pujic P., Purnelle B., Rapoport G., Ray M., Reynolds S.,  
 RA Rieger M., Rivolta C., Roche B., Roche B., Rose M., Sadate Y.,  
 RA Sato T., Scanlan E., Schleich S., Schroeder R., Scofield F.,  
 RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,  
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takenaru K.,  
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terpera P., Toononi A.,  
 RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,  
 RA Viart A., Wambut R., Wedler E., Wedler H., Weltzenegger T.,  
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasunoto K., Yata K.,  
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Zanchin A.;  
 RT "The complete genome sequence of the Gram-positive bacterium *Bacillus*  
 RL *subtilis*."; Nature 390:249-256(1997).  
 CC -1- CATALYTIC ACTIVITY: 4-imidazole-5-propanoate + H(2)O = N-  
 CC formamino-L-glutamate.  
 CC -1- PATHWAY: Histidine degradation; third step.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Potential).  
 CC -1- SIMILARITY: BELONGS TO THE HUT1 FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; D31856; BAA0642.1; -;  
 DR EMBL; Z99124; CAB15973.1; -;  
 DR Subtilist; BG1100; hut1.  
 DR InterPro; IPR002604; ATZ\_TRZ.  
 DR Pfam; PF01685; ATZ\_TRZ; 1.  
 DR TIGRPFAM; TIGR01224; hut1; 1.  
 KW Hydrolase; Histidine metabolism; Complete proteome.  
 SQ SEQUENCE 421 AA; 45564 MW; A0E1893961745771 CRC64;  
 Query Match 68.9%; Score 31; DB 1; Length 421;  
 Best Local Similarity 66.7%; Pred. No. 23;  
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 AITFNAOYA 9  
 Db 356 AITVNAAYA 364  
 RESULT 6  
 ID LEP4\_ECOLI STANDARD; PRT; 225 AA.  
 AC P25960;  
 DT 01-MAY-1992 (Rel. 22, Created)  
 DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Type 4 prepilin-like protein leader processing enzyme  
 DE [includes: Leader peptidase (EC 3.4.99.-) (Prepilin peptidase); N-  
 DE methyltransferase (EC 2.1.1.-)].  
 GN HOPE OR HOPO OR HOPO OR B3335.  
 OS *Escherichia coli*.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC Escherichia.  
 OX NCBI\_Taxid=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=95047556; PubMed=7959070;  
 RA Whitchurch C.B., Matlick J.S.;  
 RT "Escherichia coli contains a set of genes homologous to those  
 RT involved in protein secretion, DNA uptake and the assembly of type-4  
 RT fimbriae in other bacteria."  
 RL Gene 150:9-15(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MG1655;  
 RX MEDLINE=9742617; PubMed=9278503;  
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Colado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.;  
 RT "The complete genome sequence of Escherichia coli K-12."  
 RL Science 277:1453-1474(1997).  
 RN [3]  
 RP SEQUENCE OF 116-225 FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=89291745; PubMed=2661540;  
 RA Andrews S.C., Harrison P.M., Guest J.R.;  
 RT "Cloning, sequencing, and mapping of the bacterioferritin gene (bfr)  
 RT of Escherichia coli K-12."  
 RL J. Bacteriol. 171:3940-3947(1989).  
 RN [4]  
 RP IDENTIFICATION.  
 RX MEDLINE=91285432; PubMed=1676385;  
 RA Whitchurch C.B., Hobbs M., Livingston S.P., Krishnapillai V.,  
 RA Matlick J.S.;  
 RT "Characterisation of a Pseudomonas aeruginosa twitching motility gene  
 RT and evidence for a specialised protein export system widespread in  
 RT eubacteria."  
 RL Gene 101:33-44(1991).  
 CC -1- FUNCTION: CLEAVES TYPE-4 FIMBRIAL LEADER SEQUENCE AND METHYLATES  
 CC -1- THE N-TERMINAL (GENERALLY PHE) RESIDUE (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane  
 CC (Probable).  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A24.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; L28106; AAC36928.1; -  
 DR EMBL; U18997; AA58132.1; -  
 DR EMBL; AE000409; AAC76360.1; -  
 DR EMBL; M27176; AAC13988.1; -  
 DR MEROPS; A24.001; -  
 DR EcoGene; EG1359; hofD.  
 DR InterPro; IPR000045; Peptidase\_C20.  
 DR Pfam; PF01478; Peptidase\_C20; 1.  
 DR PRINTS; PR00864; PRPILNPTASE.  
 KM Hydroxylase; Protelase; Transport; Transmembrane; Inner membrane;  
 KM Transferrase; Complete proteome.  
 FT TRANSMEM 3 23 POTENTIAL.  
 FT TRANSMEM 68 88 POTENTIAL.  
 FT TRANSMEM 104 124 POTENTIAL.  
 FT TRANSMEM 128 148 POTENTIAL.  
 FT TRANSMEM 175 195 POTENTIAL.  
 FT TRANSMEM 203 223 POTENTIAL.  
 FT TRANSMEM 131 131 POTENTIAL.  
 FT TRANSMEM 131 131 A -> P (IN REF. 3).  
 FT TRANSMEM 208 209 WL -> CV (IN REF. 2; AA58132).  
 FT TRANSMEM 225 AA; 24957 MW; 90297BB7E6592EDD CRC64;  
 SQ SEQUENCE

Query Match 66.7%; Score 30; DB 1; Length 225;  
 Best Local Similarity 71.4%; Pred. No. 20;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 Oy 2 ITENA0Y 8  
 Db 113 LVFNA0Y 119  
 RESULT 7  
 RECF\_STAM STANDARD; PRT; 370 AA.  
 ID RECF\_STAM  
 AC P29232;  
 DT 01-DEC-1992 (Rel. 24, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE DNA replication and repair protein recf.  
 GN RECF OR SAV0004 OR SA0004 OR MM0004.  
 OS Staphylococcus aureus (strain N315) / ATCC 700699)  
 OS Staphylococcus aureus (strain MW2), and  
 OS Staphylococcus aureus.  
 OC Bacteria; Firmicutes; Bacillales; Staphylococcus.  
 OX NCBI\_Taxid=158878, 158879, 196620, 1280;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=M50 / ATCC 700699, and N315;  
 RX MEDLINE=21311952; PubMed=11418146;  
 RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,  
 RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Iian J.-O., Ito T.,  
 RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hoshiyama A.,  
 RA Mizumori U. Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,  
 RA Sekizima K., Hirakawa H., Kohara S., Goto S., Yabuzaki J.,  
 RA Kenehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,  
 RA Hackett M., Ogasawara N., Hayashi H., Hiramatsu K.;  
 RT "Whole genome sequencing of methicillin-resistant Staphylococcus  
 RT aureus."  
 RL Lancet 357:1225-1240(2001).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MW2;  
 RX MEDLINE=22040717; PubMed=12044378;  
 RA Baba T., Takeuchi F., Kuroda M., Yuzawa H., Aoki K.-I., Oguchi A.,  
 RA Nagai Y., Iwama N., Amano K., Naimi T., Kuroda H., Cui L.,  
 RA Yamamoto K., Hiramatsu K.;  
 RT "genome and virulence determinants of high virulence community-  
 RT acquired MRSA."  
 RL Lancet 359:1819-1827(2002).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=VR886;  
 RX MEDLINE=95206242; PubMed=7898435;  
 RA Alonso J.C., Fisher L.M.;  
 RT "Nucleotide sequence of the recF gene cluster from Staphylococcus  
 RT aureus and complementation analysis in Bacillus subtilis recF  
 RT mutants."  
 RL Mol. Gen. Genet. 246:680-686(1995).  
 RN [4]  
 RP SEQUENCE OF 267-370 FROM N.A.  
 RX MEDLINE=92165734; PubMed=1311298;  
 RA Margerison E.E.C., Howells R., Fisher L.M.;  
 RT "Nucleotide sequence of the Staphylococcus aureus gyrB-gyrA locus  
 RT encoding the DNA gyrase A and B proteins."  
 RL J. Bacteriol. 174:1596-1603(1992).  
 RN [5]  
 RP SEQUENCE OF 282-370 FROM N.A.  
 RC STRAIN=601055;  
 RX MEDLINE=93273692; PubMed=8388872;  
 RA Brookbank S.M.V., Barch P.T.;  
 RT "Cloning, sequencing, and expression of the DNA gyrase genes from  
 RT Staphylococcus aureus."  
 RL J. Bacteriol. 175:3269-3277(1993).  
 CC -1- FUNCTION: THE RECF PROTEIN IS INVOLVED IN DNA METABOLISM; IT IS



REQUIRED FOR DNA REPLICATION AND NORMAL SOS INDUCIBILITY. RECF BINDS PREFERENTIALLY TO SINGLE-STRANDED, LINEAR DNA. IT ALSO SEEMS TO BIND ATP.

-1- SUBCELLULAR LOCATION: Cytoplasmic (Potential).

-1- SIMILARITY: BELONGS TO THE RECF FAMILY.

-----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

EMBL; AF003358; BAB56166.1; -  
 EMBL; AF003359; BAB41220.1; -  
 EMBL; AF004822; BAB93869.1; -  
 EMBL; M86227; AAA73950.1; -  
 EMBL; X71437; CAA50569.1; -  
 PIR; C42295; C42295.  
 DR InterPro; IPR001338; RecF, 1.  
 DR TIGRfam; TIGR00611; recf, 1.  
 DR PROSITE; PS00617; RECF\_1; 1.  
 DR PROSITE; PS00618; RECF\_2; 1.  
 KM DNA damage; DNA replication; DNA-binding; SOS response; DNA repair;  
 NP ATP-binding; Complete proteome.  
 FT NP BIND 30 37 ATP (POTENTIAL).  
 FT CONFLICT 309 316 EXPILLD -> NIFSVC (IN REF. 4).  
 SQ SEQUENCE 370 AA; 42415 MW; 2D6471A9483B1368 CRC64;

Query Match 66.7%; Score 30; DB 1; Length 370;  
 Best Local Similarity 75.0%; Pred. No. 34;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 ITFNAOYA 9  
 Db 61 IRFNADYA 68

RESULT 8

U25\_EBV STANDARD; PRT; 570 AA.  
 AC P03333;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 01-DEC-1992 (Rel. 24, Last annotation update)  
 DE Virion protein BVRF1 (EC-RF2).  
 GN BVRF1.  
 OS Epstein-Barr virus (strain B95-8) (Human herpesvirus 4).  
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 OC Gammaherpesvirinae; Lymphocryptovirinae.  
 NCBI\_TaxID=10377;  
 RX MEDLINE=85035713; PubMed=6092825;  
 RA Bankier A.T., Dellingner P.L., Farrell P.J., Barrell B.G.,  
 RT Epstein-Barr virus.";  
 RT "Sequence analysis of the 17,166 base-pair EcoRI fragment C of B95-8  
 Mol. Biol. Med. 1:21-45(1983).  
 CC -1- FUNCTION: VIRION PROTEIN (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 U25,  
 CC EHV-1 36, EBV BVRF1, HCMV UL77, ILTV ORF2, AND VZV 34.  
 CC EHV-1 36, EBV BVRF1, HCMV UL77, ILTV ORF2, AND VZV 34.  
 CC -----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

EMBL; V01555; CAA24800.1; -  
 DR PIR; A03797; OOEERZ.

InterPro; IPR002493; U25.  
 DR Pfam; PF01499; U25; 1.  
 SQ SEQUENCE 570 AA; 62460 MW; 708D7B9B67064BAB CRC64;

Query Match 66.7%; Score 30; DB 1; Length 570;  
 Best Local Similarity 75.0%; Pred. No. 55;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 ITFNAOYA 9  
 Db 128 LTGNAOYA 135

RESULT 9

PGH2 HORSE STANDARD; PRT; 604 AA.  
 ID PGH2 HORSE  
 AC O19183;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Prostaglandin G/H synthase 2 precursor (EC 1.14.99.1) (Cyclooxygenase 2) (COX-2) (Prostaglandin-endoperoxide synthase 2) (Prostaglandin H2 synthase 2) (PGH synthase 2) (PGHS-2) (PHS II).  
 DE PTGS2 OR COX2.  
 GN Equus caballus (Horse).  
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Euteria; Perissodactyla; Equidae; Equus.  
 NCBI\_TaxID=9796;  
 RX MEDLINE=98187796; PubMed=9528947;  
 RA Boerboom D., Sirolis J.;  
 RT "Molecular characterization of equine prostaglandin G/H synthase-2  
 RT and regulation of its messenger ribonucleic acid in preovulatory  
 RT follicles.";  
 RT Endocrinology 139:1662-1670(1998).  
 CC -1- FUNCTION: MAY HAVE A ROLE AS A MAJOR MEDIATOR OF INFLAMMATION  
 CC AND/OR A ROLE FOR PROSTANOID SIGNALING IN ACTIVITY-DEPENDENT  
 CC PLASTICITY.  
 CC -1- CATALYTIC ACTIVITY: Arachidonate + AH(2) + 2 O(2) = prostaglandin  
 CC H2 + A + H(2)O.  
 CC -1- PATHWAY: FIRST STEP IN THE FORMATION OF PROSTAGLANDINS AND  
 CC THROMBOXANES.  
 CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: Membrane-associated. Microsomal membrane.  
 CC -1- MISCELLANEOUS: THIS ENZYME ACTS BOTH AS A DIOXYGENASE AND AS A  
 CC PEROXIDASE.  
 CC -1- MISCELLANEOUS: THIS ENZYME IS THE TARGET OF NONSTEROIDAL  
 CC ANTI-INFLAMMATORY DRUGS SUCH AS ASPIRIN.  
 CC -1- SIMILARITY: CONTRAINS 1 EGF-LIKE DOMAIN.  
 CC -1- SIMILARITY: BELONGS TO THE PROSTAGLANDIN G/H SYNTHASE FAMILY.  
 CC -----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

EMBL; AF027335; AAC48808.1; -  
 EMBL; AF027334; AAC07911.1; -  
 HSP; Q05769; 3PGH.  
 DR InterPro; IPR002007; Anim\_peroxidase.  
 DR InterPro; IPR000561; EGF-like.  
 DR Pfam; PF00008; EGF\_1.  
 DR Pfam; PF03098; An\_peroxidase; 1.  
 DR PRINTS; PR00457; ANPEROXIDASE.  
 DR SMART; SM00181; EGF, 1.  
 DR PROSITE; PS00022; EGF\_1; FALSE NEG.  
 DR PROSITE; PS01186; EGF\_2; FALSE NEG.  
 KM Oxioreductase; Dioxygenase; Peroxidase; Glycoprotein; Acetylation;  
 KM Prostaglandin biosynthesis; Heme; Iron; Signal; Membrane.

FT SIGNAL 1 17 BY SIMILARITY.  
 FT CHAIN 18 604 PROTAGLANDIN G/H SYNTHASE 2.  
 FT DOMAIN 18 55 EGF-LIKE.  
 FT ACT\_SITE 193 193 DISTAL HISTIDINE (BY SIMILARITY).  
 FT ACT\_SITE 371 371 CYCLOOXYGENASE (BY SIMILARITY).  
 FT BINDING 374 374 PROXIMAL HEME LIGAND (BY SIMILARITY).  
 FT DISULFID 21 32 BY SIMILARITY.  
 FT DISULFID 26 42 BY SIMILARITY.  
 FT DISULFID 44 54 BY SIMILARITY.  
 FT DISULFID 22 145 BY SIMILARITY.  
 FT DISULFID 555 561 BY SIMILARITY.  
 FT MOD\_RES 516 516 ASPARIN-ACTIVATED SERINE  
 (BY SIMILARITY).  
 FT CARBOHYD 53 53 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 130 130 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 396 396 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 580 580 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT SEQUENCE 604 AA; 68847 MW; 2879287FEB513068 CRC64;

Query Match 66.7%; Score 30; DB 1; Length 604;  
 Best Local Similarity 83.3%; Pred. No. 58;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 3 TENAOY 8  
 Db 115 TYNAOY 120

RESULT 10  
 DYHC DROME STANDARD; PRT; 4639 AA.  
 AC P37276;  
 DT 01-OCT-1994 (Rel. 30, Created)  
 DT 01-OCT-1994 (Rel. 30, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Dynein heavy chain, cytosolic (DYHC).  
 GN CDHC OR DHC64C.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;  
 OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
 OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 OC NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94375524; PubMed=8089180;  
 RA Li M., McGrail M., Serr M., Hays T.S.;  
 RT "Drosophila cytoplasmic dynein, a microtubule motor that is  
 RT asymmetrically localized in the oocyte."  
 RL J. Cell Biol. 126:1475-1494 (1994).  
 RN [2]  
 RP SEQUENCE OF 1877-1998 FROM N.A.  
 RX MEDLINE=94243034; PubMed=8186464;  
 RA Rasmussen K., Serr M., Geppert J., Gibbons I., Hays T.S.;  
 RT "A family of dynein genes in Drosophila melanogaster."  
 RL Mol. Biol. Cell 5:45-55 (1994).  
 CC -1- FUNCTION: DYNEIN HAS ATPASE ACTIVITY. CYTOPLASMIC DYNEIN ACTS AS A  
 CC MOTOR FOR THE INTRACELLULAR RETROGRADE MOTILITY OF VESICLES AND  
 CC ORGANELLES ALONG MICROTUBULES.  
 CC -1- SUBUNIT: CONSISTS OF AT LEAST TWO HEAVY CHAINS AND A NUMBER OF  
 CC INTERMEDIATE AND LIGHT CHAINS.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -1- SIMILARITY: BELONGS TO THE DYNEIN HEAVY CHAIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; L23195; AAA60323.1; -;  
 CC EMBL; L25122; AAA28492.1; -;  
 DR

DR FlyBase; FBgn0010349; Dhc64C.  
 DR InterPro; IPR004273; Dynein heavy.  
 DR Pfam; PF03028; Dynein heavy; 1.  
 KW Motor protein; Microtubules; Dynein; ATP-binding; coiled coil.  
 FT DOMAIN 530 565 COILED COIL (POTENTIAL).  
 FT DOMAIN 774 794 COILED COIL (POTENTIAL).  
 FT DOMAIN 1264 1368 COILED COIL (POTENTIAL).  
 FT DOMAIN 1999 2027 MICROTUBULE-BINDING (POTENTIAL).  
 FT DOMAIN 3189 3261 COILED COIL (POTENTIAL).  
 FT DOMAIN 3382 3478 COILED COIL (POTENTIAL).  
 FT DOMAIN 3723 3782 COILED COIL (POTENTIAL).  
 FT NP\_BIND 1895 1902 ATP (POTENTIAL).  
 FT NP\_BIND 2210 2217 ATP (POTENTIAL).  
 FT NP\_BIND 2580 2587 ATP (POTENTIAL).  
 FT NP\_BIND 2922 2929 ATP (POTENTIAL).  
 FT SEQUENCE 4639 AA; 530152 MW; 057A7D8600CCDD07E CRC64;

Query Match 66.7%; Score 30; DB 1; Length 4639;  
 Best Local Similarity 71.4%; Pred. No. 5; 5e+02;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 2 ITFNAOY 8  
 Db 2455 LTFNAOY 2461

RESULT 11  
 YUKI\_YERU STANDARD; PRT; 216 AA.  
 AC O87970;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Autoinducer synthesis protein yuki.  
 GN YUKI.  
 OS Yersinia nuckeri.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Yersinia.  
 OC NCBI\_TaxID=29486;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=1315;  
 RA Ackinson S., Throup J.P., Williams P., Stewart G.S.A.B.;  
 RT "A hierarchical quorum sensing system in Yersinia pseudotuberculosis  
 RT is involved in the regulation of motility and morphology."  
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: REQUIRED FOR THE SYNTHESIS OF AN ACYL-HSL AUTOINDUCER  
 CC THAT BIND TO YUKR AND WHICH IS INVOLVED IN THE REGULATION OF  
 CC MOTILITY AND MORPHOLOGY.  
 CC -1- SIMILARITY: BELONGS TO THE AUTOINDUCER SYNTHETASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AF079135; AAC28701.1; -;  
 DR InterPro; IPR001690; Autoind synch.  
 DR Pfam; PF00765; Autoind synch; 1.  
 DR PRINTS; PR01549; AUTOINDCRSYN.  
 DR ProDom; PD002752; Autoind synch; 1.  
 DR PROSITE; PS00949; AUTOINDUCERS\_SYNTH; 1.  
 KW Quorum sensing; Autoinducer synthesis.  
 SO SEQUENCE 216 AA; 25244 MW; 88DA934709FB40DA CRC64;

Query Match 64.4%; Score 29; DB 1; Length 216;  
 Best Local Similarity 83.3%; Pred. No. 31;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 3 TENAOY 8

Db 81 TENAQF 86

# RESULT 12

HTT1 STRGC STANDARD; PRT; 310 AA.  
 ID HTT1 STRGC  
 AC P42358;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Probable imidazolepropiomase (EC 3.5.2.7) (imidazole-5-propiomase hydrolase) (Fragment).  
 GN HTT1.  
 OS Streptococcus gordonii Challis.  
 CC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 CC Streptococcus.  
 OK NCBI\_TaxID=29390;  
 RN NCBI\_TaxID=29390;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=PK488;  
 RX MEDLINE=95012638; PubMed=7927711;  
 RA Kolander P.B., Andersen R.N., Ganeskumar N.;  
 RT "Nucleotide sequence of the Streptococcus gordonii PK488 coaggregation adhesin gene, scga, and ATP-binding cassette.";  
 RL Infect. Immun. 62:4469-4480(1994).  
 RN [2]  
 RP CONCEPTUAL TRANSLATION.  
 RA Batroch A.;  
 RL Unpublished observations (APR-1995).  
 CC -1- CATALYTIC ACTIVITY: 4-imidazole-5-propanoate + H(2)O = N-formyl-L-glutamate.  
 CC -1- PATHWAY: Histidine degradation; third step.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (potential).  
 CC -1- SIMILARITY: BELONGS TO THE HTT1 FAMILY.  
 CC -1- CAUTION: THIS IS A CONCEPTUAL TRANSLATION; A FRAMESHIFT WAS INTRODUCED IN POSITION 237 TO MAXIMIZE THE SIMILARITY WITH B.SUBTILIS HTT1.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; L15777; -; NOT ANNOTATED\_CDS.  
 DR InterPro; IPR002604; ATZ TRZ; 1.  
 DR Pfam; PF01685; ATZ TRZ; 1.  
 DR TIGRFAMs; TIGR01224; huti; 1.  
 KM Hydrolase; Histidine metabolism.  
 FT NON\_TER 1  
 SQ SEQUENCE 310 AA; 34191 MW; 40B9B41F79868004 CRC64;  
 Query Match 64.4%; Score 29; DB 1; Length 310;  
 Best Local Similarity 55.6%; Pred. No. 47;  
 Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 AITPNAOYA 9  
 Db 247 AVTINAYS 255

# RESULT 13

Y987 CAMJF STANDARD; PRT; 321 AA.  
 ID Y987 CAMJF  
 AC P45490; G9PNV2;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical protein Cj0987C.  
 GN Cj0987C.

OS Campylobacter jejuni.  
 CC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;  
 CC Campylobacter.  
 OK NCBI\_TaxID=197;  
 RN NCBI\_TaxID=197;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 43431 / TGH 9011;  
 RX MEDLINE=95247673; PubMed=7730270;  
 RA Hani E.K., Chan V.L.;  
 RT "Expression and characterization of Campylobacter jejuni benzoylglycine amidohydrolase (Hippuricase) gene in Escherichia coli.";  
 RL J. Bacteriol. 177:2396-2402(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NCIC 11168;  
 RX MEDLINE=20150912; PubMed=10688204;  
 RA Parkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C., Basham D., Chillingworth T., Davies R.M., Feltwell T., Holtroyd S., Jagsels K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W., Quail M.A., Randleam M.A., Rutherford K.M., van Vliet A.H.M., Whitehead S., Barrrell B.G.;  
 RT "The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hypervariable sequences.";  
 RL Nature 403:665-668(2000).  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
 CC -1- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION 65 ONWARD DUE TO A FRAMESHIFT.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; Z36940; CA85395.1; -;  
 DR EMBL; AL139076; CAB73243.1; -;  
 KM Hypothetical protein; Transmembrane, Complete proteome.  
 FT TRANSMEM 12 32 POTENTIAL.  
 FT TRANSMEM 52 72 POTENTIAL.  
 FT TRANSMEM 86 106 POTENTIAL.  
 FT TRANSMEM 109 129 POTENTIAL.  
 FT TRANSMEM 136 156 POTENTIAL.  
 FT TRANSMEM 168 188 POTENTIAL.  
 FT TRANSMEM 214 234 POTENTIAL.  
 FT TRANSMEM 254 274 POTENTIAL.  
 FT TRANSMEM 292 312 POTENTIAL.  
 FT CONFLICT 50 50 A -> T (IN REF. 1).  
 FT CONFLICT 55 55 S -> T (IN REF. 1).  
 SQ SEQUENCE 321 AA; 36069 MW; B4A3A7EDC32B457 CRC64;  
 Query Match 64.4%; Score 29; DB 1; Length 321;  
 Best Local Similarity 85.7%; Pred. No. 48;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 TENAQYA 9  
 Db 236 TEYAQYA 242

# RESULT 14

1108 HUMAN STANDARD; PRT; 325 AA.  
 ID 1108 HUMAN  
 AC Q08334;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Interleukin-10 receptor beta chain precursor (IL-10R-B) (IL-10R2)  
 DE (Cytokine receptor class-II CRP2-4).  
 GN IL10RB OR CRP24.  
 OS Homo sapiens (Human).

OC Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 RN NCBI\_TaxID=9606;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Fetal brain;  
 RX MEDLINE=93300510; PubMed=8314576;  
 RA Luftala G., Gardiner K., Uze G.;  
 RT "A new member of the cytokine receptor gene family maps on chromosome  
 RT 21 at less than 35 kb from IFNAR.";  
 RL Genomics 16:366-373 (1993).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96054036; PubMed=7563119;  
 RA Luftala G., McNelis M.G., Antonarakis S.E., Uze G.;  
 RT "Structure of the human CRFB4 gene: comparison with its IFNAR  
 RT neighbor";  
 RL J. Mol. Evol. 41:338-344 (1995).  
 RN [3]  
 RP CHARACTERIZATION.  
 RX MEDLINE=97459974; PubMed=9312047;  
 RA Kotenko S.V., Krause C.D., Izotova L.S., Pollack B.P., Wu W.,  
 RA Pestka S.;  
 RT "Identification and functional characterization of a second chain of  
 RT the interleukin-10 receptor complex.";  
 RL EMBO J. 16:5894-5903 (1997).  
 RN [4]  
 RP CHARACTERIZATION.  
 RX MEDLINE=20469498; PubMed=10875937;  
 RA Xie M.-H., Aggarwal S., Ho W.-H., Foster J., Zhang Z., Stinson J.,  
 RA Wood W.I., Goddard A.D., Garney A.L.;  
 RT "Interleukin (IL)-22, a novel human cytokine that signals through the  
 RT interferon receptor-related protein CRF2-4 and IL-22R.";  
 RL J. Biol. Chem. 275:31335-31339 (2000).  
 CC -1- FUNCTION: RECEPTOR FOR IL-10 AND IL-22. SERVES AS AN ACCESSORY  
 CC CHAIN ESSENTIAL FOR THE ACTIVE IL-10 RECEPTOR COMPLEX AND TO  
 CC INITIATE IL-10-INDUCED SIGNAL TRANSDUCTION EVENTS.  
 CC -1- SUBCELLULAR LOCATION: TYPE I membrane protein.  
 CC -1- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.  
 CC -1- SIMILARITY: BELONGS TO THE CLASS II CYTOKINE FAMILY OF RECEPTORS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; Z17227; CAAT8933.1; -;  
 DR EMBL; U08988; AAA86872.1; -;  
 DR PIR; A47003; A47003.  
 DR HSSP; P13726; 1TFH.  
 DR Genew; HANC:5965; IL10RB.  
 DR MIM; 123889; -;  
 DR InterPro; IPR000282; Cytok\_receptor\_2.  
 DR InterPro; IPR001187; Tissue\_factor.  
 DR Pfam; PF01108; Tissue\_fac\_1.  
 KM Receptor; Transmembrane; Glycoprotein; Signal.  
 FT SIGNAL 1 19  
 FT CHAIN 20 325  
 FT DOMAIN 20 220  
 FT TRANSMEM 221 242  
 FT DOMAIN 243 325  
 FT DOMAIN 113 205  
 FT DISULFID 66 74  
 FT DISULFID 188 209  
 FT CARBOHYD 49 68  
 FT CARBOHYD 68 102  
 FT CARBOHYD 102 161  
 FT CARBOHYD 161 161  
 FT CONFLICT 124 124  
 FT CONFLICT 269 273

FT CONFLICT 274 325 MISSING (IN REF. 2).  
 SQ SEQUENCE 325 AA; 37011 MW; 66706C79F8514B23 CRC64;  
 Query Match 64.4%; Score 29; DB 1; Length 325;  
 Best Local Similarity 71.4%; Pred. No. 49;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 2 ITPNAY 8  
 Db 50 LTFPAY 56  
 RESULT 15  
 ID MOUSE STANDARD; PRT; 349 AA.  
 AC 061150;  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-OCT-2001 (Rel. 40, Last sequence update)  
 DE Interleukin-10 receptor beta chain precursor (IL-10R-B) (IL-10R2)  
 DE (Cytokine receptor class-II CRF2-4).  
 GN IL10RB OR CRFB4.  
 OS Mus musculus (Mouse).  
 OC Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97199375; PubMed=9047351;  
 RA Gibbs V.C., Pennica D.;  
 RT "CRF2-4: Isolation of cDNA clones encoding the human and mouse  
 RT proteins.";  
 RL Gene 186:97-101 (1997).  
 RN [2]  
 RP CHARACTERIZATION.  
 RX MEDLINE=98130620; PubMed=9463407;  
 RA Spencer S.D., Di Marco F., Hookey J., Pitts-Week S., Bauer M.,  
 RA Ryan A.M., Soedel B., Gibbs V.C., Aguet M.;  
 RT "The orphan receptor CRF2-4 is an essential subunit of the interleukin  
 RT 10 receptor.";  
 RL J. Exp. Med. 187:571-578 (1998).  
 CC -1- FUNCTION: RECEPTOR FOR IL-10 AND IL-22. SERVES AS AN ACCESSORY  
 CC CHAIN ESSENTIAL FOR THE ACTIVE IL-10 RECEPTOR COMPLEX AND TO  
 CC INITIATE IL-10-INDUCED SIGNAL TRANSDUCTION EVENTS.  
 CC -1- SUBCELLULAR LOCATION: TYPE I membrane protein.  
 CC -1- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.  
 CC -1- SIMILARITY: BELONGS TO THE CLASS II CYTOKINE FAMILY OF RECEPTORS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; U53696; AAC53062.1; -;  
 DR MGD; MGI:109380; IL10RB.  
 DR InterPro; IPR000282; Cytok\_receptor\_2.  
 DR InterPro; IPR003961; FN\_III.  
 DR SMART; SM00060; FN3; 1.  
 KM Receptor; Transmembrane; Glycoprotein; Signal.  
 FT SIGNAL 1 19  
 FT CHAIN 20 349  
 FT DOMAIN 20 220  
 FT TRANSMEM 221 241  
 FT DOMAIN 242 349  
 FT DOMAIN 113 205  
 FT DISULFID 66 74  
 FT DISULFID 188 209  
 FT CARBOHYD 49 49  
 FT CARBOHYD 102 102  
 FT CARBOHYD 161 161  
 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FLAGP -> VERME (IN REF. 2).

FT CARBOHYD 199 199 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 349 AA; 39774 MW; 58BA4F6B8630A39 CRC64;

Query Match 64.4%; Score 29; DB 1; Length 349;

Best Local Similarity 71.4%; Pred. No. 53;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 ITFNAQY 8

Db 50 LTFTAQY 56

Search completed: January 29, 2003, 14:00:28  
 Job time : 7.23077 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2003, 13:56:21 ; Search time 25.1538 seconds  
(without alignments)  
73.723 Million cell updates/sec

Title: US-09-807-949a-138  
Perfect score: 45  
Sequence: 1 AITFNAQYA 9

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

SPTREMBL\_21:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriaph:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Length	DB ID	Description
1	34	75.6	255 16 Q9S275	Q9S275 streptomyce
2	33	73.3	238 16 Q9L1A9	Q9L1A9 streptomyce
3	33	73.3	335 16 Q9KCT8	Q9KCT8 raietonia s
4	33	73.3	335 16 Q8KST9	Q8KST9 raietonia s
5	33	73.3	462 16 Q9YB4	Q9YB4 streptococc
6	33	73.3	462 16 Q9TMM3	Q9TMM3 streptococc
7	32	71.1	250 16 Q8ZGP9	Q8ZGP9 yersinia pe
8	32	71.1	307 16 Q9ZVC7	Q9ZVC7 rhizobium m
9	32	71.1	356 16 Q97K43	Q97K43 clostridium
10	32	71.1	451 10 Q9LGE2	Q9LGE2 oryza sativ
11	32	71.1	483 16 Q9CFI1	Q9CFI1 lactococcus
12	32	71.1	520 16 Q99WY8	Q99WY8 staphylococ
13	32	71.1	525 2 Q93SM5	Q93SM5 staphylococ
14	32	71.1	587 5 Q9U7F4	Q9U7F4 onchocerca
15	32	71.1	592 16 Q9SX94	Q9SX94 staphylococ
16	32	71.1	767 16 Q9Z2Y4	Q9Z2Y4 rhizobium m

17	32	71.1	867 3 Q59780	Q59780 schizosacch
18	32	71.1	1017 2 Q9ZAG0	Q9ZAG0 porphyromon
19	32	71.1	1357 12 Q89328	Q89328 rice ragged
20	31	68.9	114 12 Q91ZB8	Q91ZB8 cercopithec
21	31	68.9	169 3 Q12550	Q12550 aspergillus
22	31	68.9	211 3 Q12549	Q12549 aspergillus
23	31	68.9	254 5 Q8WSR7	Q8WSR7 dictyostell
24	31	68.9	260 5 Q93790	Q93790 caenorhabdi
25	31	68.9	335 17 Q97YR4	Q97YR4 sulfobolus
26	31	68.9	376 4 Q969Y9	Q969Y9 homo sapien
27	31	68.9	379 4 Q96P06	Q96P06 homo sapien
28	31	68.9	383 4 Q13880	Q13880 homo sapien
29	31	68.9	383 6 Q8WN70	Q8WN70 saquinus oe
30	31	68.9	383 6 Q8WN69	Q8WN69 cercopithec
31	31	68.9	383 11 Q8VH1	Q8VH1 mus musculu
32	31	68.9	415 4 Q9KXK7	Q9KXK7 homo sapien
33	31	68.9	442 13 Q9PVJ9	Q9PVJ9 gallus gall
34	31	68.9	569 13 Q9YHW0	Q9YHW0 gallus gall
35	31	68.9	2233 16 Q9KGT0	Q9KGT0 streptococ
36	31	68.9	3187 5 Q9BLV5	Q9BLV5 leishmania
37	30	66.7	152 16 Q31554	Q31554 bacillus su
38	30	66.7	154 16 Q9KXW9	Q9KXW9 streptomyce
39	30	66.7	175 16 Q9P82	Q9P82 xyella fas
40	30	66.7	215 16 Q9KZP3	Q9KZP3 streptomyce
41	30	66.7	239 10 Q9SZG5	Q9SZG5 arabidopsis
42	30	66.7	253 16 Q25894	Q25894 helicobacte
43	30	66.7	253 16 Q9Z1P9	Q9Z1P9 helicobacte
44	30	66.7	303 16 Q97MW5	Q97MW5 clostridium
45	30	66.7	312 16 Q9KGT0	Q9KGT0 bacillus ha

## ALIGNMENTS

RESULT 1  
Q9S275 PRELIMINARY; PRT; 255 AA.  
ID Q9S275  
AC Q9S275;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE Putative enoyl-(acyl-carrier-protein) reductase.  
OS INHA OR SCO1814 OR SCI28.08C.  
OS Streptomyces coelicolor.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
OX NCBI\_TaxId=1902;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2);  
RA Seeger K., Harris D.;  
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2);  
RA Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;  
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2) / M145;  
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,  
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kleier H.,  
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,  
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,

RA Huang C.-H., Kieseer T., Larke L., Murphy L., Oliver K., O'Neil S.,  
 RA Rabbittowitsch E., Rajandream M.A., Rutherford K., Rutter S.,  
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,  
 RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,  
 RA Hopwood D.A.;  
 RT "Complete genome sequence of the model actinomycete Streptomyces  
 coelicolor A3(2).";  
 RL Nature 417:141-147(2002).  
 CC -1 - SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES  
 (SDB) FAMILY.  
 DR EMBL; AL096844; CAB50882.1; -.  
 DR HSSP; P46533; 1ENY.  
 DR InterPro; IPR002198; ADH\_short.  
 DR Pfam; PF00106; adh\_short; 1.  
 KW Oxidoreductase.  
 SQ SEQUENCE 255 AA; 27153 MW; 86FC52A7E02CBE64 CRC64;

Query Match 75.6%; Score 34; DB 16; Length 255;  
 Best Local Similarity 75.0%; Pred. No. 26;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 ITFNOYA 9  
 Db 143 LTFDOYA 150

## RESULT 2

AC Q9LIA9 PRELIMINARY; PRT; 238 AA.  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Putative secreted protein.  
 GN SC01572 OR SCL24.08.  
 OS Streptomyces coelicolor.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.  
 OX NCBI\_TaxId=1502;

RA [1]  
 RA SEQUENCE FROM N.A.  
 RC STRAIN=A3(2) / M145;  
 RA Bercley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,  
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieseer H.,  
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,  
 RA Cronin A., Frazer A., Goble A., Hidalgo J., Hornsby T., Howarth S.,  
 RA Huang C.-H., Kieseer T., Larke L., Murphy L., Oliver K., O'Neil S.,  
 RA Rabbittowitsch E., Rajandream M.A., Rutherford K., Rutter S.,  
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,  
 RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,  
 RA Hopwood D.A.;  
 RT "Complete genome sequence of the model actinomycete Streptomyces  
 coelicolor A3(2).";  
 RL Nature 417:141-147(2002).  
 DR EMBL; AL157956; CAB76092.1; -.  
 SQ SEQUENCE 238 AA; 25090 MW; 1FSA150B87E7BA7A CRC64;

Query Match 73.3%; Score 33; DB 16; Length 238;  
 Best Local Similarity 75.0%; Pred. No. 39;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 ITFNOYA 9  
 Db 161 ITFNOYA 168

## RESULT 3

AC Q9KGT8 PRELIMINARY; PRT; 335 AA.  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)

DE Flagellar motor switch.  
 GN FLIM.  
 OS Ralstonia solanacearum (Pseudomonas solanacearum).  
 OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;  
 OC Ralstonia.  
 OX NCBI\_TaxId=305;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K60;  
 RA MEDLINE=21264371; PubMed=11371523;  
 RA Tans-Kersten J., Huang H., Allen C.;  
 RT "Ralstonia solanacearum needs motility for invasive virulence on  
 tomato.";  
 RL J. Bacteriol. 183:3597-3605(2001).  
 DR EMBL; AF283286; AAF87587.1; -.  
 DR InterPro; IPR001689; flag\_FLIM.  
 DR InterPro; IPR001543; SpOA.  
 DR Pfam; PF02154; FLIM; 1.  
 DR Pfam; PF01052; SpOA; 1.  
 DR PRINTS; PR00955; FLAGMOTORFLIM.  
 SQ SEQUENCE 335 AA; 37963 MW; 9AF1B54E6E4B9BF CRC64;

Query Match 73.3%; Score 33; DB 2; Length 335;  
 Best Local Similarity 85.7%; Pred. No. 57;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 TFNOYA 9  
 Db 307 TFNOYA 313

## RESULT 4

AC Q8XST9 PRELIMINARY; PRT; 335 AA.  
 DT 01-MAR-2002 (TREMBlrel. 20, Created)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Probable flagellar motor switch protein FLIM.  
 GN FLIM OR RSP0378 OR RS00806.  
 OS Ralstonia solanacearum (Pseudomonas solanacearum).  
 OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;  
 OC Ralstonia.  
 OX NCBI\_TaxId=305;

RA [1]  
 RA SEQUENCE FROM N.A.  
 RC STRAIN=GM11000;  
 RX MEDLINE=21681879; PubMed=11623852;  
 RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,  
 RA Atilat M., Billault A., Brottier P., Camus J.C., Catolico L.,  
 RA Chandler M., Cholene N., Claudel-Renard C., Cunac S., Demange N.,  
 RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T.,  
 RA Sguiter P., Thebaud P., Whalen M., Wincker P., Levy M.,  
 RA Weisenbach J., Boucher C.A.;  
 RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";  
 RL Nature 415:497-502(2002).  
 DR EMBL; AL646078; CAD17529.1; -.  
 DR InterPro; IPR001689; flag\_FLIM.  
 DR InterPro; IPR001543; SpOA.  
 DR Pfam; PF02154; FLIM; 1.  
 DR Pfam; PF01052; SpOA; 1.  
 DR PRINTS; PR00955; FLAGMOTORFLIM.  
 KW Plasmid; Complete proteome.  
 SQ SEQUENCE 335 AA; 38027 MW; D2237199075444A5 CRC64;

Query Match 73.3%; Score 33; DB 16; Length 335;  
 Best Local Similarity 85.7%; Pred. No. 57;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 TFNOYA 9  
 Db 307 TFNOYA 313



```

RESULT 5
ID Q99YB4 PRELIMINARY; PRT; 462 AA.
AC Q99YB4;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical protein SPY1781.
GN SPY1781.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SF370 / ATCC 700294 / SEROTYPE M1;
RA MEDLINE=21192684; PubMed=11296296;
RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,
RA Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,
RA Qian X., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,
RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;
RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).
DR EMBL; AB006605; AKK4516.1; -.
DR InterPro; IPR001454; Hlg_nase/hydlase.
DR InterPro; IPR000150; Hypothet_cof.
DR Pfam; PF00702; Hydrolase; 1.
DR TIGRfams; TIGR00099; Hypothet_cof; 2.
DR PROSITE; PS01228; COF_1; UNKNOWN_1.
DR PROSITE; PS01229; COF_2; 1.
DR KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 462 AA; 51980 MW; 21A09B8A8E276ED5 CRC64;

Query Match
Best Local Similarity 73.3%; Score 33; DB 16; Length 462;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATFNQY 8
DB 62 AVTYNGQY 69

RESULT 6
ID Q97NM3 PRELIMINARY; PRT; 462 AA.
AC Q97NM3;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Cof family protein.
GN SPI997.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PIGR4;
RA MEDLINE=21357209; PubMed=11463916;
RA Tettein H., Nelson K.B., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S., Heideberg J., Deboy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gilm M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA Holzapple E., Khouri H., Wolf A.M., Uteback T.R., Hansen C.L.,
RA McDonald L.A., Feldlyum T.V., Anguoli S., Dickenson T., Hickey B.K.,
RA Holt I.B., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus
RT pneumoniae.";
RL Science 293:498-506(2001).
DR EMBL; AB007488; AAK76064.1; -.

```

```

DR TIGR; SPI997; -.
DR InterPro; IPR001454; Hlg_nase/hydlase.
DR InterPro; IPR000150; Hypothet_cof.
DR Pfam; PF00702; Hydrolase; 1.
DR TIGRfams; TIGR00099; Hypothet_cof; 2.
DR PROSITE; PS01228; COF_1; UNKNOWN_1.
DR PROSITE; PS01229; COF_2; 1.
DR KW Complete proteome.
SQ SEQUENCE 462 AA; 51898 MW; 9BAE9064182AEC97 CRC64;

Query Match
Best Local Similarity 73.3%; Score 33; DB 16; Length 462;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATFNQY 8
DB 62 AVTYNGQY 69

RESULT 7
ID Q8ZGP9 PRELIMINARY; PRT; 250 AA.
AC Q8ZGP9;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DE 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Hypothetical protein YPO1238.
GN YPO1238.
OS Yersinia pestis.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Yersinia.
OX NCBI_TaxID=632;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CO-92 / BIOVAR ORIENTALIS;
RA MEDLINE=21470413; PubMed=11586360;
RA Parhill J., Wren B.W., Thomson N.R., Tibball R.W., Holden M.T.G.,
RA Prentice M.B., Sebahia M., James K.D., Churcher C., Mungall K.L.,
RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdano-Tarraga A.M.,
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA Felwell T., Hamlin N., Holtroyd S., Jagels K., Karlyshev A.V.,
RA Leather S., Moul S., Oyston P.C.F., Quail M., Rutherford K.,
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrett B.G.;
RT "Genome sequence of Yersinia pestis, the causative agent of plague.";
RL Nature 413:523-527(2001).
DR EMBL; AJ414147; CAC90072.1; -.
DR KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 250 AA; 28573 MW; 061B1DCFB81CB7F1B CRC64;

Query Match
Best Local Similarity 71.1%; Score 32; DB 16; Length 250;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATFNQYA 9
DB 98 SITFDEQYA 106

RESULT 8
ID Q92VC7 PRELIMINARY; PRT; 307 AA.
AC Q92VC7;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Putative inosine-uridine preferring nucleoside hydrolase
DE protein.
GN RB0781 OR SMR21277.
OS Rhizobium meliloti (sinorhizobium meliloti).
OG Plasmid pSymB (megaplasmid 2).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;

```

RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=1021;  
 RX MEDLINE=21396508; PubMed=11481431;  
 RA Pinan T.M., Weidner S., Wong K., Buhrmaster J., Chain P.,  
 RA Vorholster F.U., Hernandez-Lucas I., Becker A., Cowie A., Gouzy J.,  
 RA Golding B., Puelher A.;  
 RT "The complete sequence of the 1,683-kb *PSymB* megaplasmid from the N2-  
 RT fixing endosymbiont *Sinorhizobium meliloti*."  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9889-9894(2001).  
 DR EMBL; AF603644; CAC49181.1; -  
 DR InterPro; IPR001910; I/U\_NHdase.  
 DR Pfam; PF01156; I/U\_Nuc\_hydro; 1.  
 DR ProDom; PD007736; I/U\_NHdase; 1.  
 KM Hydrolyase; Plasmid; Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 307 AA; 32201 MW; 63A8B47DC4DF0257 CRC64;

Query Match 71.1%; Score 32; DB 16; Length 307;  
 Best Local Similarity 85.7%; Pred. No. 85;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 TFNAOYA 9  
 DB 279 TFNAHYA 285

## RESULT 9

ID Q97K43 PRELIMINARY; PRT; 356 AA.  
 AC Q97K43;  
 DT 01-OCT-2001 (TREMBlrel. 18, Created)  
 DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Predicted phosphohydrolyase, ICC family.  
 GN CAC1078.  
 OS Clostridium acetobutylicum.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;  
 OC Clostridiales; Clostridiaceae; Clostridium.  
 NCBI\_TaxID=148;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;  
 RX MEDLINE=21359325; PubMed=11466286;  
 RA Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,  
 RA Gibson R., Lee H.M., Dubois J., Olu D., Hittl J., Wolf Y.I.,  
 RA Tegusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,  
 RA Bennett G.N., Koonin E.V., Smith D.R.;  
 RT "Genome sequence and comparative analysis of the solvent-producing  
 RT bacterium *Clostridium acetobutylicum*."  
 RL J. Bacteriol. 183:4823-4838(2001).  
 DR EMBL; AB007623; AAK79052.1; -  
 DR InterPro; IPR004843; M-dpstrase.  
 DR InterPro; IPR004844; S/R\_phosphatase.  
 DR Pfam; PF00149; Metallophos; 1.  
 KM Hydrolyase; Complete proteome.  
 SQ SEQUENCE 356 AA; 39168 MW; EBA28A27BE734ED CRC64;

Query Match 71.1%; Score 32; DB 16; Length 356;  
 Best Local Similarity 66.7%; Pred. No. 1e+02;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 AITENAOYA 9  
 DB 25 SITNAOYS 33

## RESULT 10

ID Q9LGE2 PRELIMINARY; PRT; 451 AA.  
 AC Q9LGE2;  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)

DE Putative zinc finger protein.  
 GN P0462H08.6.  
 OS Oryza sativa (Rice).  
 OC Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Eriophytaceae; Oryzaceae; Oryza.  
 NCBI\_TaxID=4530;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. NIPPONBARE;  
 RA Sasaki T., Matsumoto T., Yamamoto K.;  
 RT "Oryza sativa nippobare(GA3) genomic DNA, chromosome 1, PAC  
 RT clone, P0462H08."  
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AP002525; BAB07983.1; -  
 DR InterPro; IPR000571; Znf\_CCH.  
 DR Pfam; PF00642; Zf-CCH; 5.  
 DR SMART; SM00356; Znf\_C3H1; 5.  
 SQ SEQUENCE 451 AA; 47997 MW; 3BE8F98AC8AED181 CRC64;

Query Match 71.1%; Score 32; DB 10; Length 451;  
 Best Local Similarity 55.6%; Pred. No. 1.3e+02;  
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 AITENAOYA 9  
 DB 244 AVTYGSOYA 252

## RESULT 11

ID Q9CF11 PRELIMINARY; PRT; 483 AA.  
 AC Q9CF11;  
 DT 01-JUN-2001 (TREMBlrel. 17, Created)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Sugar ABC transporter substrate binding protein.  
 GN YPCG OR LL1495.  
 OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;  
 OC Streptococcaceae; Lactococcus.  
 NCBI\_TaxID=1360;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=IL1403;  
 RX MEDLINE=21235186; PubMed=11337471;  
 RA Bolotin A., Wincker P., Mauer S., Jaillon O., Malame K.,  
 RA Weissenbach J., Ehrlich S.D., Sorokin A.;  
 RT "The complete genome sequence of the lactic acid bacterium *Lactococcus*  
 RT lactis ssp. lactis IL1403."  
 RL Genome Res. 11:731-753(2001).  
 DR EMBL; AB006380; AAK05593.1; -  
 DR InterPro; IPR000567; SBP\_bac\_1.  
 DR Pfam; PF01547; SBP\_bacterial\_1; 1.  
 KM Complete proteome.  
 SQ SEQUENCE 483 AA; 52938 MW; 858F5A20BD8BA2DA CRC64;

Query Match 71.1%; Score 32; DB 16; Length 483;  
 Best Local Similarity 85.7%; Pred. No. 1.4e+02;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 ITFNKOY 8  
 DB 157 ITFNKOY 163

## RESULT 12

ID Q99WY8 PRELIMINARY; PRT; 520 AA.  
 AC Q99WY8;  
 DT 01-JUN-2001 (TREMBlrel. 17, Created)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)

DE Hypothetical protein SAV0235.  
 GN SAV0235 OR SA0227.  
 OS Staphylococcus aureus (strain Mu50 / ATCC 700699), and  
 OC Staphylococcus aureus (strain N315).  
 CC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;  
 OC Staphylococcus.  
 RN NCBI\_TaxId=158879; 158879;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=S. aureus (strain Mu50), and S. aureus (strain N315);  
 RX MEDLINE=21311952; PubMed=11418146;  
 RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,  
 Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,  
 Kanemori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,  
 Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,  
 Sekimizu K., Hiramatsu H., Kohara S., Goto S., Yabuzaki J.,  
 Kanesaka M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,  
 Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;  
 RT "Whole genome sequencing of methicillin-resistant Staphylococcus  
 aureus.";  
 RL Lancet 357:1225-1240(2001).  
 DR EMBL; AP003358; BAB56397.1; -;  
 DR EMBL; AP003129; BAB41449.1; -;  
 DR InterPro; IPR004165; CoA\_trans.1.  
 DR Pfam; PF01144; CoA\_trans.1.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 520 AA; 58357 MW; 059353593D4B5D9 CRC64;

Query Match 71.1%; Score 32; DB 16; Length 520;  
 Best Local Similarity 75.0%; Pred. No. 1.5e+02;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 ITFNAOYA 9  
 Db 439 IDFNAOYS 446

RESULT 13

093SM5 PRELIMINARY; PRT; 525 AA.  
 AC 093SM5;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
 DE Putative acetyl-CoA:acetoacetyl-CoA transferase Fadd.  
 GN FADD.  
 OS Staphylococcus aureus.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;  
 OC Staphylococcus.  
 RN NCBI\_TaxId=1280;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Longshaw C.M.;  
 RT "Identification of a fatty acid degradation operon in Staphylococcus  
 aureus.";  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF033081; AAK5154.1; -;  
 DR InterPro; IPR004165; CoA\_trans.  
 DR Pfam; PF01144; CoA\_trans.1.  
 KW Transferase.  
 SQ SEQUENCE 525 AA; 58876 MW; 0EC4BFCC425D7B78 CRC64;

Query Match 71.1%; Score 32; DB 2; Length 525;  
 Best Local Similarity 75.0%; Pred. No. 1.5e+02;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 ITFNAOYA 9  
 Db 444 IDFNAOYS 451

RESULT 14  
 09U7F4

ID 09U7F4 PRELIMINARY; PRT; 587 AA.  
 AC 09U7F4;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE P-glycoprotein-like protein.  
 GN PLP1.  
 OS Onchocerca volvulus.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;  
 OC Onchocercidae; Onchocerca.  
 RN NCBI\_TaxId=6282;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=9426471; PubMed=10498183;  
 RA Huang Y.-J., Pichard R.K.;  
 RT "Identification and stage-specific expression of two putative P-  
 glycoprotein coding genes in Onchocerca volvulus.";  
 RL Mol. Biochem. Parasitol. 102:273-281(1999).  
 CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.  
 DR EMBL; AF128532; AAD49563.1; -;  
 DR InterPro; IPR003593; AAA\_ATPase.  
 DR InterPro; IPR001140; ABCTransporter.  
 DR InterPro; IPR003439; ABCTransporter.  
 DR Pfam; PF00664; ABC\_trans.1.  
 DR Pfam; PF000005; ABC\_trans.1.  
 DR Prodom; PD000006; ABC\_trans.1.  
 DR SMART; SM00382; AAA; 1.  
 DR PROSITE; PS00211; ABC\_TRANSPORTER; 1.  
 KW ATP-binding; Transport.  
 SQ SEQUENCE 587 AA; 65676 MW; 8ED0B84B6A0D6F9B CRC64;

Query Match 71.1%; Score 32; DB 5; Length 587;  
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 FNAOYA 9  
 Db 243 FNAOYA 248

RESULT 15

099X94 PRELIMINARY; PRT; 592 AA.  
 AC 099X94;  
 DT 01-JUN-2001 (TREMBLrel. 17, Created)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
 DE Hypothetical protein SAV0121.  
 GN SAV0121 OR SA0117.  
 OS Staphylococcus aureus (strain Mu50 / ATCC 700699), and  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;  
 OC Staphylococcus.  
 RN NCBI\_TaxId=158879; 158879;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=S. aureus (strain Mu50), and S. aureus (strain N315);  
 RX MEDLINE=21311952; PubMed=11418146;  
 RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,  
 Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,  
 Kanemori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,  
 Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,  
 Sekimizu K., Hiramatsu H., Kohara S., Goto S., Yabuzaki J.,  
 Kanesaka M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,  
 Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;  
 RT "Whole genome sequencing of methicillin-resistant Staphylococcus  
 aureus.";  
 RL Lancet 357:1225-1240(2001).  
 DR EMBL; AP003358; BAB56283.1; -;  
 DR EMBL; AP003129; BAB41336.1; -;  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 592 AA; 69050 MW; 8910CDC420FB769B CRC64;

Query Match 71.1%; Score 32; DB 16; Length 592;  
 Best Local Similarity 77.8%; Pred. No. 1.7e+02;  
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AITPNAOYA 9  
 |||||  
 Db 371 AIPFNALYA 379

Search completed: January 29, 2003, 14:02:27  
 Job time : 26.1538 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2003, 13:57:26 ; Search time 11.0769 Seconds  
(without alignments)  
23.906 Million cell updates/sec

Title: US-09-807-949A-138

Perfect score: 45

Sequence: 1 AITFNAQYA 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA:\*  
1: /cgn2\_6/prodata/1/1aa/5A\_COMB.pep:\*  
2: /cgn2\_6/prodata/1/1aa/5B\_COMB.pep:\*  
3: /cgn2\_6/prodata/1/1aa/5A\_COMB.pep:\*  
4: /cgn2\_6/prodata/1/1aa/5B\_COMB.pep:\*  
5: /cgn2\_6/prodata/1/1aa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/prodata/1/1aa/Backfill1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	31	68.9	184 1	US-08-044-621D-30
2	31	68.9	184 1	US-08-044-621D-31
3	31	68.9	184 1	US-08-044-621D-31
4	31	68.9	184 2	US-08-047-370-1
5	31	68.9	185 1	US-08-709-912-2
6	31	68.9	185 2	US-08-709-912-2
7	31	68.9	185 2	US-09-047-370-2
8	31	68.9	185 4	US-09-230-590-2
9	31	68.9	211 1	US-07-842-349-2
10	31	68.9	211 1	US-08-244-686-2
11	31	68.9	722 4	US-08-961-083-84
12	29	64.4	202 5	PCT-US94-1427-3
13	29	64.4	224 4	US-08-871-572B-11
14	29	64.4	223 4	US-08-871-572B-8
15	29	64.4	325 2	US-08-683-743-4
16	29	64.4	333 4	US-09-267-031-4
17	28	62.2	186 2	US-08-833-610-3
18	28	62.2	186 3	US-08-834-033A-13
19	28	62.2	491 4	US-09-314-268-3
20	28	62.2	523 4	US-09-550-338-2
21	27	60.0	541 4	US-09-134-001C-4481
22	27	60.0	31 2	US-08-468-790-18
23	27	60.0	137 4	US-09-456-830-15
24	27	60.0	137 4	US-09-456-830-15
25	27	60.0	137 4	US-09-456-830-19
26	27	60.0	137 4	US-09-456-830-17
27	27	60.0	137 4	US-09-002-285-7
27	27	60.0	137 4	US-09-002-285-15

28	27	60.0	137 4	US-09-002-285-19	Sequence 19, Appl
29	27	60.0	137 4	US-09-002-285-37	Sequence 37, Appl
30	27	60.0	157 1	US-08-257-999-4	Sequence 4, Appl
31	27	60.0	191 4	US-08-936-165A-525	Sequence 525, App
32	27	60.0	199 3	US-08-737-248-9	Sequence 9, Appl
33	27	60.0	226 2	US-08-658-277-4	Sequence 4, Appl
34	27	60.0	313 3	US-08-732-412-2	Sequence 21, Appl
35	27	60.0	354 4	US-08-970-264A-21	Sequence 2, Appl
36	27	60.0	444 4	US-08-044-718-1	Sequence 1, Appl
37	27	60.0	454 2	US-08-491-835-25	Sequence 26, Appl
38	27	60.0	454 3	US-08-946-092A-26	Sequence 26, Appl
39	27	60.0	454 4	US-09-172-062-26	Sequence 26, Appl
40	27	60.0	454 4	US-09-301-520D-26	Sequence 26, Appl
41	27	60.0	454 5	PCT-US94-00685-26	Sequence 36, Appl
42	27	60.0	465 4	US-09-071-035-36	Sequence 36, Appl
43	27	60.0	467 1	US-08-151-574-32	Sequence 32, Appl
44	27	60.0	467 1	US-08-146-424-20	Sequence 20, Appl
45	27	60.0	467 1	US-08-693-709-2	Sequence 2, Appl

#### ALIGNMENTS

RESULT 1  
US-08-044-621D-30  
; Sequence 30, Application US/08044621D  
; Patent No. 5405769  
; GENERAL INFORMATION:  
; APPLICANT: Warren W. Makarchuk  
; APPLICANT: Wang L. Sung  
; APPLICANT: Makoto Yaguchi  
; APPLICANT: Robert L. Campbell  
; APPLICANT: David R. Rose  
; TITLE OF INVENTION: CONSTRUCTION OF THERMOSTABLE MUTANTS  
; TITLE OF INVENTION: OF A LOW MOLECULAR MASS XYLANASE  
; NUMBER OF SEQUENCES: 37  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Gowing, Streeby & Henderson  
; STREET: Suite 2600, 160 Elgin Street  
; CITY: Ottawa  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: K1P 1G3  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 5.25 in., 360KB storage  
; COMPUTER: IBM PC  
; OPERATING SYSTEM: PC-DOS  
; SOFTWARE: WordPerfect 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/044,621D  
; FILING DATE: April 8, 1993  
; CLASSIFICATION: 435  
; PRIOR APPLICATION NUMBER:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Judy A. Erratic  
; REGISTRATION NUMBER: 34,076  
; REFERENCE/DOCKET NUMBER: 08-863796  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 613-786-0199  
; TELEFAX: 613-563-9869  
; TELETYPE:  
; INFORMATION FOR SEQ ID NO: 30:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 184  
; TYPE: Amino Acid  
; STRANDEDNESS: No. 5405769 Relevant  
; TOPOLOGY: linear  
; MOLECULE TYPE:  
; DESCRIPTION: protein  
; HYPOTHETICAL: No  
; ANTI-SENSE: No

FRAGMENT TYPE: NO  
ORIGINAL SOURCE:  
ORGANISM: *Aspergillus niger*, var. *awamori*  
IMMEDIATE SOURCE:  
POSITION IN GENOME:  
FEATURE:  
PUBLICATION INFORMATION:  
AUTHORS: Maat, J., Roza, M., Verbakel, J., Stam,  
AUTHORS: H., Santos da Silva, M.J., Egmund, M.R.,  
AUTHORS: Hagemans, M.L.D., Gorcom, R.F.M.V.,  
AUTHORS: Heesing, J.G.M., Hondel, C.A.M.J.O.V.d.  
AUTHORS: & Rotterdam, C.V.  
TITLE:  
JOURNAL: *Xylans and Xylanases*  
VOLUME:  
ISSUE:  
PAGES: 349-360  
DATE: 1992  
DOCUMENT NUMBER:  
FILING DATE:  
PUBLICATION DATE:  
RELEVANT RESIDUES IN SEQ ID NO:  
US-08-044-621D-30

Query Match 68.9%; Score 31; DB 1; Length 184;  
Best Local Similarity 55.6%; Pred. No. 28;  
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 AITPNAQYA 9  
Db 51 AITYSABYS 59

RESULT 2  
US-08-044-621D-31  
Sequence 31, Application US/08044621D  
Patent No. 5405769  
GENERAL INFORMATION:  
APPLICANT: Warren W. Wakarchuk  
APPLICANT: Ming L. Sung  
APPLICANT: Makoto Yaguchi  
APPLICANT: Robert L. Campbell  
APPLICANT: David R. Rose  
TITLE OF INVENTION: CONSTRUCTION OF THERMOSTABLE MUTANTS  
TITLE OF INVENTION: OF A LOW MOLECULAR MASS XYLANASE  
NUMBER OF SEQUENCES: 37  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Gowling, Strachy & Henderson  
STREET: Suite 2600, 160 Elgin Street  
CITY: Ottawa  
STATE: Ontario  
COUNTRY: Canada  
ZIP: K1P 1C3  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 5.25 in., 360KB storage  
COMPUTER: IBM PC  
OPERATING SYSTEM: PC-DOS  
SOFTWARE: Wordperfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/044,621D  
FILING DATE: April 8, 1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Judy A. Erratt  
REGISTRATION NUMBER: 34,076  
REFERENCE/DOCKET NUMBER: 08-863796  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 613-786-0199  
TELEFAX: 613-563-9869  
TELEX:

INFORMATION FOR SEQ ID NO: 31:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 184  
TYPE: Amino Acid  
STRANDEDNESS: NO. 5405769 Relevant  
TOPOLOGY: linear  
MOLECULE TYPE:  
DESCRIPTION: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: NO  
ORIGINAL SOURCE:  
ORGANISM: *Aspergillus tubigenis*  
STRAIN: *Aspergillus tubigenis* A  
IMMEDIATE SOURCE:  
POSITION IN GENOME:  
FEATURE:  
PUBLICATION INFORMATION:  
AUTHORS: de Graaff, L.H., van den Broeck, H.C.,  
AUTHORS: van Ooijf, A.J.J. & Visser, J.  
TITLE:  
JOURNAL: *Xylans and Xylanases*  
VOLUME:  
ISSUE:  
PAGES: 235-246  
DATE: 1992  
DOCUMENT NUMBER:  
FILING DATE:  
PUBLICATION DATE:  
RELEVANT RESIDUES IN SEQ ID NO:  
US-08-044-621D-31

Query Match 68.9%; Score 31; DB 1; Length 184;  
Best Local Similarity 55.6%; Pred. No. 28;  
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 AITPNAQYA 9  
Db 51 AITYSABYS 59

RESULT 3  
US-08-709-912-1  
Sequence 1, Application US/08709912  
Patent No. 5759840  
GENERAL INFORMATION:  
APPLICANT: Sung Dr., Ming L  
APPLICANT: Yaguchi Dr., Makoto  
APPLICANT: Ishikawa Dr., Kazuhiko  
TITLE OF INVENTION: Modification of Xylanase to improve  
TITLE OF INVENTION: Thermophilicity, Alkalophilicity and  
TITLE OF INVENTION: Thermostability  
NUMBER OF SEQUENCES: 54  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Fitzpatrick, Cella, Harper, and Scinto  
STREET: 277 Park Ave.  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10172-0194  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/709,912  
FILING DATE: 09-SEP-1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Olsen Mr, Warren E  
REGISTRATION NUMBER: 27290  
REFERENCE/DOCKET NUMBER: 1039.2000

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 758-2400  
TELEFAX: (212) 758-2982  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 184 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: Internal  
ORIGINAL SOURCE:  
ORGANISM: *Aspergillus niger*, var. *awamori*  
PUBLICATION INFORMATION:  
AUTHORS: Moat Dr, J  
AUTHORS: Dr Roga, M  
AUTHORS: Stam, H  
AUTHORS: Santos da Silva, M J  
AUTHORS: Egmond, M R  
AUTHORS: Hagemans, M.L. D  
AUTHORS: Gorcom, R.F.M.V.  
AUTHORS: Hessing, J.G.M.  
AUTHORS: Hondel, C.A.M.U  
JOURNAL: *Xylan and Xylanase*  
PAGES: 349-360  
DATE: 1992  
US-08-709-912-1

Query Match  
Best Local Similarity 55.6%; Score 31; DB 1; Length 184;  
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATENAOYA 9  
|||:|:|:  
Db 51 AITSAAYS 59

RESULT 4  
US-09-047-370-1  
; Sequence 1, Application US/09047370  
; Patent No. 5866408  
; GENERAL INFORMATION:  
; APPLICANT: Sung Dr., Wing L  
; APPLICANT: Yaguchi Dr., Makoto  
; APPLICANT: Ishikawa Dr., Kazuhiko  
; TITLE OF INVENTION: Modification of Xylanase to Improve  
; TITLE OF INVENTION: Thermophilicity, Alkalophilicity and  
; TITLE OF INVENTION: Thermostability  
; NUMBER OF SEQUENCES: 54  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fitzpatrick, Cella, Harper, and Scinto  
; STREET: 277 Park Ave.  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10172-0194  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/047,370  
; FILING DATE:  
; CLASSIFICATION:  
; PRIORITY APPLICATION DATA:  
; APPLICATION NUMBER: US 08/709,912  
; FILING DATE: 09-SEP-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Olsen Mr, Warren E

REGISTRATION NUMBER: 27290  
REFERENCE/DOCKET NUMBER: 1039.2000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 758-2400  
TELEFAX: (212) 758-2982  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 184 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: Internal  
ORIGINAL SOURCE:  
ORGANISM: *Aspergillus niger*, var. *awamori*  
PUBLICATION INFORMATION:  
AUTHORS: Moat Dr, J  
AUTHORS: Dr Roga, M  
AUTHORS: Stam, H  
AUTHORS: Santos da Silva, M J  
AUTHORS: Egmond, M R  
AUTHORS: Hagemans, M.L. D  
AUTHORS: Gorcom, R.F.M.V.  
AUTHORS: Hessing, J.G.M.  
AUTHORS: Hondel, C.A.M.U  
JOURNAL: *Xylan and Xylanase*  
PAGES: 349-360  
DATE: 1992  
US-09-047-370-1

Query Match  
Best Local Similarity 55.6%; Score 31; DB 2; Length 184;  
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATENAOYA 9  
|||:|:|:  
Db 51 AITSAAYS 59

RESULT 5  
US-08-709-912-2  
; Sequence 2, Application US/08709912  
; Patent No. 5759840  
; GENERAL INFORMATION:  
; APPLICANT: Sung Dr., Wing L  
; APPLICANT: Yaguchi Dr., Makoto  
; APPLICANT: Ishikawa Dr., Kazuhiko  
; TITLE OF INVENTION: Modification of Xylanase to Improve  
; TITLE OF INVENTION: Thermophilicity, Alkalophilicity and  
; TITLE OF INVENTION: Thermostability  
; NUMBER OF SEQUENCES: 54  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fitzpatrick, Cella, Harper, and Scinto  
; STREET: 277 Park Ave.  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10172-0194  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/709,912  
; FILING DATE: 09-SEP-1996  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Olsen Mr, Warren E  
; REGISTRATION NUMBER: 27290

REFERENCE/DOCKET NUMBER: 1039.2000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 758-2400  
TELEFAX: (212) 758-2982  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 185 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHEITICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: internal  
ORIGINAL SOURCE:  
ORGANISM: Aspergillus tubigenensis  
PUBLICATION INFORMATION:  
AUTHORS: de Graaf, L.H.  
AUTHORS: van der Broeck, H.C.  
AUTHORS: van Ooijck, A.J.J.  
AUTHORS: Visser, J  
JOURNAL: Xylan and Xylanase  
PAGES: 235-246  
DATE: 1992  
US-08-709-912-2

Query Match 68.9%; Score 31; DB 1; Length 185;  
Best Local Similarity 55.6%; Pred. No. 28;  
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 AITENAOYA 9  
|||::|||:  
Db 52 AITSAEYS 60

RESULT 6  
US-09-047-370-2  
Sequence 2, Application US/09047370  
Patent No. 5866408  
GENERAL INFORMATION:  
APPLICANT: Sung Dr., Wing L  
APPLICANT: Yaguchi Dr., Makoto  
TITLE OF INVENTION: Modification of Xylanase to Improve  
TITLE OF INVENTION: Thermophilicity, Alkalophilicity and  
TITLE OF INVENTION: Thermostability  
NUMBER OF SEQUENCES: 54  
CORRESPONDENCE ADDRESS:  
ADDRESS: Fitzpatrick, Cella, Harper, and Scinto  
STREET: 277 Park Ave.  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10172-0194  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/047,370  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/709,912  
FILING DATE: 09-SEP-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Olsen Mr, Warren E  
REGISTRATION NUMBER: 27290  
REFERENCE/DOCKET NUMBER: 1039.2000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 758-2400  
TELEFAX: (212) 758-2982

INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 185 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHEITICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: internal  
ORIGINAL SOURCE:  
ORGANISM: Aspergillus tubigenensis  
PUBLICATION INFORMATION:  
AUTHORS: de Graaf, L.H.  
AUTHORS: van der Broeck, H.C.  
AUTHORS: van Ooijck, A.J.J.  
AUTHORS: Visser, J  
JOURNAL: Xylan and Xylanase  
PAGES: 235-246  
DATE: 1992  
US-09-047-370-2

Query Match 68.9%; Score 31; DB 2; Length 185;  
Best Local Similarity 55.6%; Pred. No. 28;  
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 AITENAOYA 9  
|||::|||:  
Db 52 AITSAEYS 60

RESULT 7  
US-09-230-590-2  
Sequence 2, Application US/09230590  
Patent No. 6361808  
GENERAL INFORMATION:  
APPLICANT: SOUPE, Jerome  
TITLE OF INVENTION: BEUDEKER, Robert Franciscus  
TITLE OF INVENTION: PROCESS FOR THE PRODUCTION OF ALCOHOLIC BEVERAGES  
TITLE OF INVENTION: USING MALTSEED  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESS: LADAS & PARRY  
STREET: 26 WEST 61 STREET  
CITY: NEW YORK  
STATE: NY  
COUNTRY: USA  
ZIP: 10023  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: Microsoft Windows for workgroups  
SOFTWARE: Wordperfect 8  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/230,590  
FILING DATE: 28-APR-1999  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/EP97/04016  
FILING DATE: 23-JUL-1997  
APPLICATION NUMBER: EP 96202195.2  
FILING DATE: 05-AUG-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: MASS, CLIFFORD J.  
REGISTRATION NUMBER: 30,086  
REFERENCE/DOCKET NUMBER: U-012094-9  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 708-1890  
TELEFAX: (212) 246-8959  
TELEX: 233288  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 185 amino acids



TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 2:  
US-09-230-590-2

Query Match 68.9%; Score 31; DB 4; Length 185;  
Best Local Similarity 55.6%; Pred. No. 28;  
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AITENAOYA 9  
|||:|:|:  
Db 52 AITYSAYS 60

RESULT 8  
US-07-842-349-2  
; Sequence 2, Application US/07842349  
; Patent No. 535864

GENERAL INFORMATION:  
APPLICANT: VAN DEN BROECK, HENRIETTE C.  
APPLICANT: DE GRAAF, LEENDERT H.  
APPLICANT: HILDE R., JAN D.  
APPLICANT: VAN OUYEN J., ALBERT J.  
APPLICANT: VISSER, JACOB  
APPLICANT: HARDER, ABRAHAM  
TITLE OF INVENTION: CLONING AND EXPRESSION OF XYLANASE GNEBS  
TITLE OF INVENTION: FROM FUNGAL ORIGIN  
NUMBER OF SEQUENCES: 23  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: MORRISON & FORSTER  
STREET: 755 Page Mill Road  
CITY: Palo Alto  
STATE: California  
COUNTRY: USA  
ZIP: 94304-1018  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/842,349  
FILING DATE: 19920427

CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: MURASHIGE, KATE H.  
REGISTRATION NUMBER: 29,959  
REFERENCE/DOCKET NUMBER: 24615-20031.00  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 813-5600  
TELEFAX: (415) 494-0792  
TELEX: 706141  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 211 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-07-842-349-2

Query Match 68.9%; Score 31; DB 1; Length 211;  
Best Local Similarity 55.6%; Pred. No. 32;  
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AITENAOYA 9  
|||:|:|:  
Db 78 AITYSAYS 86

RESULT 9  
US-08-244-686-2  
; Sequence 2, Application US/08244686

Patent No. 5705358  
GENERAL INFORMATION:  
APPLICANT:  
TITLE OF INVENTION: Process for the production of a protein signals  
TITLE OF INVENTION: using endoxylanase II (exII) expression signals  
NUMBER OF SEQUENCES: 12  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/244,686  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 211 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-244-686-2

Query Match 68.9%; Score 31; DB 1; Length 211;  
Best Local Similarity 55.6%; Pred. No. 32;  
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AITENAOYA 9  
|||:|:|:  
Db 78 AITYSAYS 86

RESULT 10  
US-08-961-083-84  
; Sequence 84, Application US/08961083  
; Patent No. 6159469  
GENERAL INFORMATION:  
APPLICANT: Choi et. al.  
TITLE OF INVENTION: Streptococcus pneumoniae Antigenes and Vaccines  
NUMBER OF SEQUENCES: 452  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/961,083  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Brookes, A. Anders  
REGISTRATION NUMBER: 36,373  
REFERENCE/DOCKET NUMBER: PB340P2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512  
INFORMATION FOR SEQ ID NO: 84:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 722 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-961-083-84

Query Match 68.9%; Score 31; DB 4; Length 722;  
Best Local Similarity 62.5%; Pred. No. 1.3e+02;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 AITFMAOY 8  
DB 557 AVTLNABY 564

## RESULT 11

PCT-US94-14277-3  
Sequence 3, Application PC/TUS9414277  
GENERAL INFORMATION:  
APPLICANT: Aguet, Michel  
APPLICANT: Bohni, Ruth  
APPLICANT: Hemmi, Silvio  
TITLE OF INVENTION: Receptor Subunit Polypeptides  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 460 Point San Bruno Blvd  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patin (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US94/14277  
FILING DATE: 07-DEC-1994  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/164596  
FILING DATE: 09-DEC-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Love, Richard B.  
REGISTRATION NUMBER: 34,659  
REFERENCE/DOCKET NUMBER: 866PCT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415/225-5530  
TELEFAX: 415/952-9881  
TELEX: 910/371-7168  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 202 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
PCT-US94-14277-3

Query Match 64.4%; Score 29; DB 5; Length 202;  
Best Local Similarity 50.0%; Pred. No. 81;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 AITFMAOY 8  
DB 31 SVTFSABY 38

## RESULT 12

US-08-871-572B-11  
Sequence 11, Application US/08871572B  
Patent No. 6287853  
GENERAL INFORMATION:  
APPLICANT: Pestka, Sidney  
APPLICANT: Kotenko, Serguei  
APPLICANT: Soh, Jaemog  
APPLICANT: Donnelly, Robert  
APPLICANT: Mariano, Thomas  
APPLICANT: Cook, Jeffrey  
APPLICANT: Emmanuel, Stuart

APPLICANT: Schwartz, Barbara  
TITLE OF INVENTION: Accessory Factor for Interferon Gamma  
TITLE OF INVENTION: and Its Receptor  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Richard R. Muccino  
STREET: 758 Springfield Avenue  
CITY: Summit  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07901

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Releasee #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/871,572B  
FILING DATE: 9-JUNE-1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Muccino, Richard R.  
REGISTRATION NUMBER: 32,538  
REFERENCE/DOCKET NUMBER: UMD1-011  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (908) 273-4679  
TELEFAX: (908) 273-4988  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 224 amino acids  
TYPE: amino acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-871-572B-11

Query Match 64.4%; Score 29; DB 4; Length 224;  
Best Local Similarity 50.0%; Pred. No. 91;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 AITFMAOY 8  
DB 57 SVTFSABY 64

## RESULT 13

US-08-871-572B-8  
Sequence 8, Application US/08871572B  
Patent No. 6287853  
GENERAL INFORMATION:  
APPLICANT: Pestka, Sidney  
APPLICANT: Kotenko, Serguei  
APPLICANT: Soh, Jaemog  
APPLICANT: Donnelly, Robert  
APPLICANT: Mariano, Thomas  
APPLICANT: Cook, Jeffrey  
APPLICANT: Emmanuel, Stuart  
TITLE OF INVENTION: Accessory Factor for Interferon Gamma  
TITLE OF INVENTION: and Its Receptor  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Richard R. Muccino  
STREET: 758 Springfield Avenue  
CITY: Summit  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07901  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible

```
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patentin Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/871,572B
/ FILING DATE: 9-JUNE-1997
/ CLASSIFICATION: 435
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Muccino, Richard R.
/ REGISTRATION NUMBER: 32,538
/ REFERENCE/DOCKET NUMBER: UMD1-011
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (908) 273-4988
/ TELEFAX: (908) 273-4679
/ INFORMATION FOR SEQ ID NO: 8:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 233 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: unknown
/ TOPOLOGY: unknown
/ MOLECULE TYPE: peptide
/ HYPOTHETICAL: NO
/ ANTI-SENSE: NO
/
US-08-871-572B-8

Query Match          64.4%; Score 29; DB 4; Length 233;
Best Local Similarity 71.4%; Pred. No. 95;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 ITFNAQY 8
   :|||
Db 50 LTFNAQY 56

RESULT 14
US-08-683-743-4
/ Sequence 4, Application US/08683743
/ Patent No. 5843697
/ GENERAL INFORMATION:
/ APPLICANT: Pestka, Sidney
/ APPLICANT: Kotenko, Serguei
/ TITLE OF INVENTION: CYTOKINE RECEPTOR SIGNAL TRANSDUCTION
/ TITLE OF INVENTION: CHAIN
/ NUMBER OF SEQUENCES: 25
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: David A. Jackson, Esq.
/ STREET: 411 Hackensack Ave, Continental Plaza, 4th
/ STREET: Floor
/ CITY: Hackensack
/ STATE: New Jersey
/ COUNTRY: USA
/ ZIP: 07601
/
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patentin Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/683,743
/ FILING DATE: 17-JUL-1996
/ CLASSIFICATION: 435
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Jackson Esq., David A.
/ REGISTRATION NUMBER: 26,742
/ REFERENCE/DOCKET NUMBER: 601-1-050
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 201-343-1684
/ TELEFAX: 201-343-1684
/ INFORMATION FOR SEQ ID NO: 4:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 325 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
```

```
/ MOLECULE TYPE: Protein
/ HYPOTHETICAL: NO
/ FRAGMENT TYPE:
/
US-08-683-743-4

Query Match          64.4%; Score 29; DB 2; Length 325;
Best Local Similarity 71.4%; Pred. No. 1.4e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 ITFNAQY 8
   :|||
Db 50 LTFNAQY 56

RESULT 15
US-09-267-031-4
/ Sequence 4, Application US/09267031
/ Patent No. 6137031
/ GENERAL INFORMATION:
/ APPLICANT: Zhang, Yuelin
/ APPLICANT: Kinkema, Mark
/ APPLICANT: Dong, Xinlian
/ APPLICANT: Ronald, Pamela
/ APPLICANT: Chern, MawSheng
/ APPLICANT: The Regents of the University of California
/ TITLE OF INVENTION: DNA Binding Proteins That Interact With NPPI
/ FILE REFERENCE: 023070-092500US
/ CURRENT APPLICATION NUMBER: US/09/267,031
/ CURRENT FILING DATE: 1999-03-11
/ NUMBER OF SEQ ID NOS: 26
/ SOFTWARE: Patentin Ver. 2.0
/ SEQ ID NO 4
/ LENGTH: 333
/ TYPE: PRT
/ ORGANISM: Oryzias sp.
/
US-09-267-031-4

Query Match          64.4%; Score 29; DB 4; Length 333;
Best Local Similarity 55.6%; Pred. No. 1.4e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 ITFNAQYA 9
   :|||
Db 114 AMTFDEYA 122
```

Search completed: January 29, 2003, 14:04:21  
Job time : 12.0769 secs

**THIS PAGE BLANK (uspto)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2003, 14:00:01 ; Search time 6.92308 Seconds  
(without alignments)  
26.232 Million cell updates/sec

Title: US-09-807-949A-138  
Perfect score: 45  
Sequence: 1 AITFNAQYA 9

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 122226 seqs, 20178551 residues

Total number of hits satisfying chosen parameters: 122226

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

Published Applications AA.\*  
1: /cgn2\_6/ptodata/1/pubppa/US08\_NEW\_PUB.pep.\*  
2: /cgn2\_6/ptodata/1/pubppa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubppa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubppa/US07\_NEW\_PUB.pep.\*  
5: /cgn2\_6/ptodata/1/pubppa/US07\_PUBCOMB.pep.\*  
6: /cgn2\_6/ptodata/1/pubppa/US07\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubppa/US07\_PUBCOMB.pep.\*  
8: /cgn2\_6/ptodata/1/pubppa/US09\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubppa/US09\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/1/pubppa/US09\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubppa/US10\_NEW\_PUB.pep.\*  
12: /cgn2\_6/ptodata/1/pubppa/US10\_PUBCOMB.pep.\*  
13: /cgn2\_6/ptodata/1/pubppa/US60\_NEW\_PUB.pep.\*  
14: /cgn2\_6/ptodata/1/pubppa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	31	68.9	185	9	US-09-970-616-2
2	31	68.9	722	9	US-09-765-272-84
3	30	66.7	256	9	US-09-885-913A-182
4	30	66.7	547	10	US-09-961-527A-11
5	29	64.4	40	10	US-09-864-761-388A9
6	29	64.4	199	9	US-09-728-911-35
7	29	64.4	325	9	US-10-066-500-137
8	29	64.4	325	9	US-10-174-590-390
9	29	64.4	325	9	US-10-176-758-390
10	29	64.4	325	9	US-10-175-737-390
11	29	64.4	325	10	US-09-870-574-3
12	29	64.4	325	10	US-09-949-192-5
13	29	64.4	325	12	US-10-052-586-390
14	29	64.4	437	10	US-09-925-297-684
15	29	64.4	1564	10	US-09-801-368-244
16	28	62.2	236	10	US-09-881-752A-340
17	28	62.2	279	9	US-09-738-626-6899
18	28	62.2	356	9	US-09-738-626-4781
19	28	62.2	523	9	US-09-738-626-4177

20	28	62.2	523	10	US-09-767-878-2	Sequence 2, Appli
21	28	62.2	550	10	US-09-841-132-562	Sequence 562, App
22	28	62.2	633	9	US-09-834-998A-3	Sequence 3, Appli
23	28	62.2	921	9	US-09-117-447-6	Sequence 6, Appli
24	28	62.2	1827	10	US-09-819-247-2	Sequence 2, Appli
25	27	60.0	191	10	US-09-939-980-525	Sequence 525, App
26	27	60.0	218	10	US-09-825-561A-6	Sequence 6, Appli
27	27	60.0	231	10	US-09-925-299-874	Sequence 874, App
28	27	60.0	245	9	US-09-738-626-4876	Sequence 4876, Ap
29	27	60.0	289	9	US-09-941-973-2	Sequence 2, Appli
30	27	60.0	341	10	US-09-815-242-10657	Sequence 10657, A
31	27	60.0	376	9	US-09-931-457A-49	Sequence 49, Appl
32	27	60.0	377	9	US-09-931-457A-47	Sequence 47, Appl
33	27	60.0	438	10	US-09-430-221-3	Sequence 3, Appli
34	27	60.0	454	12	US-10-115-406-26	Sequence 26, Appl
35	27	60.0	455	9	US-09-813-399-39	Sequence 39, Appl
36	27	60.0	467	12	US-10-083-452-11	Sequence 11, Appl
37	27	60.0	471	10	US-09-815-242-10445	Sequence 10445, A
38	27	60.0	471	10	US-09-815-242-14028	Sequence 14028, A
39	27	60.0	474	10	US-09-752-639-148	Sequence 148, App
40	27	60.0	474	10	US-09-964-198-148	Sequence 148, App
41	27	60.0	478	9	US-09-943-123-6	Sequence 6, Appli
42	27	60.0	478	9	US-09-943-123-7	Sequence 7, Appli
43	27	60.0	478	9	US-09-943-123-12	Sequence 12, Appl
44	27	60.0	478	9	US-09-943-123-13	Sequence 13, Appl
45	27	60.0	478	10	US-09-813-555-1	Sequence 1, Appli

## ALIGNMENTS

RESULT 1  
US-09-970-616-2  
Sequence 2, Application US/09970616  
Patent No. US20020164399A1  
GENERAL INFORMATION:  
APPLICANT: SOUPPE, Jerome  
BENDEKER, Robert Franciscus  
TITLE OF INVENTION: PROCESS FOR THE PRODUCTION OF ALCOHOLIC BEVERAGES USING MAL  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSER: LADAS & PARRY  
STREET: 26 WEST 61 STREET  
CITY: <Unknown>  
STATE:  
COUNTRY: USA  
ZIP CODE: 10023  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: Microsoft Windows for workgroups  
SOFTWARE: WordPerfect 8  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/970,616  
FILING DATE: 04-Oct-2001  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/230,590  
FILING DATE: 28-APR-1999  
APPLICATION NUMBER: PCT/EP97/04016  
FILING DATE: 23-JUL-1997  
APPLICATION NUMBER: EP 96202195.2  
ATTORNEY/AGENT INFORMATION:  
FILING DATE: 05-AUG-1996  
NAME: MASS, CLIFFORD J.  
REGISTRATION NUMBER:  
TELECOMMUNICATION INFORMATION: (C) REF./DOCKET NO.: U-012094-9  
TELEPHONE: 233288  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:

LENGTH: 185 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 SEQUENCE DESCRIPTION: SEQ ID NO: 2:  
 US-09-970-616-2

Query Match 68.9%; Score 31; DB 9; Length 185;  
 Best Local Similarity 55.6%; Pred. No. 14;  
 Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 AITFNAQYA 9  
 |||:|:|:  
 Db 52 AITFSAEYS 60

RESULT 2  
 US-09-765-272-84

Sequence 84, Application US/09765272  
 Patent No. US20020061545A1  
 GENERAL INFORMATION:

APPLICANT: Choi et. al.  
 TITLE OF INVENTION: Streptococcus pneumoniae Antigenes and Vaccines  
 NUMBER OF SEQUENCES: 452  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Human Genome Sciences, Inc.  
 STREET: 9410 Key West Avenue  
 CITY: Rockville  
 STATE: Maryland  
 COUNTRY: USA  
 ZIP: 20850

# COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
 COMPUTER: HP Vectra 486/33  
 OPERATING SYSTEM: MSDOS version 6.2  
 SOFTWARE: ASCII Text

# CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/765,272  
 FILING DATE: 22-Jan-2001  
 CLASSIFICATION: <Unknown>  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 08/961,083  
 FILING DATE: <Unknown>

# ATTORNEY/AGENT INFORMATION:

NAME: Brookes, A. Anders  
 REGISTRATION NUMBER: 36,373  
 REFERENCE/DOCKET NUMBER: PB340P2  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (301) 309-8504  
 TELEFAX: (301) 309-8512

# INFORMATION FOR SEQ ID NO: 84:

SEQUENCE CHARACTERISTICS:  
 LENGTH: 722 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 SEQUENCE DESCRIPTION: SEQ ID NO: 84:  
 US-09-765-272-84

Query Match 68.9%; Score 31; DB 10; Length 722;  
 Best Local Similarity 62.5%; Pred. No. 63;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 AITFNAQY 8  
 |||:|:|:  
 Db 557 AITLNAEY 564

# RESULT 3

US-09-895-913A-182  
 Sequence 182, Application US/09895913A  
 Patent No. US20020160456A1

GENERAL INFORMATION:  
 APPLICANT: Kleantous, Harold  
 APPLICANT: Al-Garawi, Amal  
 APPLICANT: Miller, Charles  
 APPLICANT: Tomb, Jean Francois  
 APPLICANT: Oomen, Raymond P.  
 TITLE OF INVENTION: Identification of Polynucleotides  
 Encoding No. US20020160456A1 Helicobacter Polypeptides in the f  
 TITLE OF INVENTION: Genome  
 FILE REFERENCE: 06133/043002  
 CURRENT APPLICATION NUMBER: US/09/895,913A  
 CURRENT FILING DATE: 2001-06-29  
 PRIOR APPLICATION NUMBER: US 08/881,227  
 PRIOR FILING DATE: 1997-06-24  
 NUMBER OF SEQ ID NOS: 368  
 SOFTWARE: FastSeq for Windows Version 4.0  
 SEQ ID NO 182  
 LENGTH: 256  
 TYPE: PRT  
 ORGANISM: Helicobacter pylori  
 US-09-895-913A-182

Query Match 66.7%; Score 30; DB 9; Length 256;  
 Best Local Similarity 62.5%; Pred. No. 33;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 ITFNAQYA 9  
 |||:|:|:  
 Db 111 LTFGAETA 118

# RESULT 4

US-09-961-527A-11  
 Sequence 11, Application US/09961527A  
 Patent No. US2002014232A1  
 GENERAL INFORMATION:

APPLICANT: Wang, Xun  
 APPLICANT: Turgeon, B. Gillian  
 APPLICANT: Yoder, Olen  
 APPLICANT: Wu, Jiansuo  
 TITLE OF INVENTION: Fungal target genes and methods to identify those genes  
 FILE REFERENCE: TM0129-0T  
 CURRENT APPLICATION NUMBER: US/09/961,527A  
 CURRENT FILING DATE: 2001-09-24  
 PRIOR APPLICATION NUMBER: US 60/234,673  
 PRIOR FILING DATE: 2000-09-22  
 PRIOR APPLICATION NUMBER: US 60/234,650  
 PRIOR FILING DATE: 2000-09-22  
 NUMBER OF SEQ ID NOS: 19  
 SEQ ID NO 11  
 LENGTH: 547  
 TYPE: PRT  
 ORGANISM: Cochliobolus  
 US-09-961-527A-11

Query Match 66.7%; Score 30; DB 10; Length 547;  
 Best Local Similarity 71.4%; Pred. No. 76;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 3 TFNAQYA 9  
 |||:|:|:  
 Db 425 TFNAQFS 431

# RESULT 5

US-09-864-761-38849  
 Sequence 38849, Application US/09864761  
 Patent No. US20020048763A1  
 GENERAL INFORMATION:

APPLICANT: Penn, Sharon G.  
 APPLICANT: Rank, David R.  
 APPLICANT: Hanzel, David K.  
 APPLICANT: Chen, Wensheng

TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
FILE REFERENCE: Aecmca-X-1  
CURRENT APPLICATION NUMBER: US/09/864,761  
CURRENT FILING DATE: 2001-05-23  
PRIORITY APPLICATION NUMBER: US 60/180,312  
PRIORITY FILING DATE: 2000-02-04  
PRIORITY APPLICATION NUMBER: US 60/207,456  
PRIORITY FILING DATE: 2000-05-26  
PRIORITY APPLICATION NUMBER: US 09/632,366  
PRIORITY FILING DATE: 2000-06-03  
PRIORITY APPLICATION NUMBER: GB 24263.6  
PRIORITY FILING DATE: 2000-10-04  
PRIORITY APPLICATION NUMBER: US 60/236,359  
PRIORITY FILING DATE: 2000-09-27  
PRIORITY APPLICATION NUMBER: PCT/US01/00666  
PRIORITY FILING DATE: 2001-01-30  
PRIORITY APPLICATION NUMBER: PCT/US01/00667  
PRIORITY FILING DATE: 2001-01-30  
PRIORITY APPLICATION NUMBER: PCT/US01/00664  
PRIORITY FILING DATE: 2001-01-30  
PRIORITY APPLICATION NUMBER: PCT/US01/00669  
PRIORITY FILING DATE: 2001-01-30  
PRIORITY APPLICATION NUMBER: PCT/US01/00665  
PRIORITY FILING DATE: 2001-01-30  
PRIORITY APPLICATION NUMBER: PCT/US01/00668  
PRIORITY FILING DATE: 2001-01-30  
PRIORITY APPLICATION NUMBER: PCT/US01/00663  
PRIORITY FILING DATE: 2001-01-30  
PRIORITY APPLICATION NUMBER: PCT/US01/00670  
PRIORITY FILING DATE: 2001-01-30  
PRIORITY APPLICATION NUMBER: US 60/234,687  
PRIORITY FILING DATE: 2000-09-21  
PRIORITY APPLICATION NUMBER: US 09/608,408  
PRIORITY FILING DATE: 2000-06-30  
PRIORITY APPLICATION NUMBER: US 09/774,203  
PRIORITY FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 38849  
LENGTH: 40  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AP000111.1  
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 5  
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 4.2  
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.4  
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.5  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 9.3  
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 4.1  
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4.6  
OTHER INFORMATION: SWISSPROT HIT: Q08334, EVALUATE 3.00e-18  
OTHER INFORMATION: EST\_HUMAN HIT: AA307672.1, EVALUATE 4.00e-17  
US-09-864-761-38849  
Query Match 64.4%; Score 29; DB 10; Length 40;  
Best Local Similarity 71.4%; Pred. No. 7;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Patent No. US20020012669A1  
GENERAL INFORMATION:  
APPLICANT: Preenell, Scott R.  
APPLICANT: Xu, Wenteng  
APPLICANT: Kinsvogel, Wayne  
APPLICANT: Chen, Zhi  
TITLE OF INVENTION: Human Cytokine Receptor  
FILE REFERENCE: 99-93  
CURRENT APPLICATION NUMBER: US/09/728,911  
CURRENT FILING DATE: 2000-12-01  
PRIORITY APPLICATION NUMBER: US 60/169,049  
PRIORITY FILING DATE: 1999-12-03  
PRIORITY APPLICATION NUMBER: US 60/232,219  
PRIORITY FILING DATE: 2000-09-13  
PRIORITY APPLICATION NUMBER: US 60/244,610  
PRIORITY FILING DATE: 2000-10-31  
NUMBER OF SEQ ID NOS: 36  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 35  
LENGTH: 199  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-728-911-35  
Query Match 64.4%; Score 29; DB 10; Length 199;  
Best Local Similarity 71.4%; Pred. No. 40;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
QY 2 ITFNAQY 8  
DB 31 LTFNAQY 37  
RESULT 7  
US-10-066-500-137  
Sequence 137, Application US/10066500  
Patent No. US20020177165A1  
GENERAL INFORMATION:  
APPLICANT: Avi J. Ashkenazi  
APPLICANT: Kevin P. Baker  
APPLICANT: David A. Botstein  
APPLICANT: Luc Desnoyers  
APPLICANT: Dan L. Baton  
APPLICANT: Napoleone Ferrara  
APPLICANT: Sherman Fong  
APPLICANT: Wei-Qiang Gao  
APPLICANT: Hanspeter Gerber  
APPLICANT: Mary E. Gerritsen  
APPLICANT: Audrey Goddard  
APPLICANT: Paul J. Godowski  
APPLICANT: Austin L. Gurney  
APPLICANT: Ivar J. Kjaava  
APPLICANT: Jennie P. Mather  
APPLICANT: Mary A. Napier  
APPLICANT: James Pan  
APPLICANT: Nicholas F. Paoni  
APPLICANT: Margaret Ann Roy  
APPLICANT: Timothy A. Stewart  
APPLICANT: Daniel Tumes  
APPLICANT: Colin K. Watanabe  
APPLICANT: P. Mickey Williams  
APPLICANT: William I. Wood  
APPLICANT: Zemin Zang  
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
FILE REFERENCE: P3130R1C7  
CURRENT APPLICATION NUMBER: US/10/066,500  
CURRENT FILING DATE: 2002-02-01  
PRIORITY APPLICATION NUMBER: 10/002,796  
PRIORITY FILING DATE: 2001-11-15  
PRIORITY APPLICATION NUMBER: 60/056974  
PRIORITY FILING DATE: 1997-08-26  
PRIORITY APPLICATION NUMBER: 60/059115

```

PRIOR FILING DATE: 1997-09-17
PRIOR APPLICATION NUMBER: 60/059263
PRIOR FILING DATE: 1997-09-18
PRIOR APPLICATION NUMBER: 60/059588
PRIOR FILING DATE: 1997-09-17
PRIOR APPLICATION NUMBER: 60/062285
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/062816
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: 60/063082
PRIOR FILING DATE: 1997-10-31
PRIOR APPLICATION NUMBER: 60/063329
PRIOR FILING DATE: 1997-10-27
PRIOR APPLICATION NUMBER: 60/063733
PRIOR FILING DATE: 1997-10-29
PRIOR APPLICATION NUMBER: 60/066364
PRIOR FILING DATE: 1997-11-21
PRIOR APPLICATION NUMBER: 60/066840
PRIOR FILING DATE: 1997-11-25
PRIOR APPLICATION NUMBER: 60/069694
PRIOR FILING DATE: 1997-12-16
PRIOR APPLICATION NUMBER: 60/074086
PRIOR FILING DATE: 1998-02-09
PRIOR APPLICATION NUMBER: 60/074092
PRIOR FILING DATE: 1998-02-09
PRIOR APPLICATION NUMBER: 60/079294
PRIOR FILING DATE: 1998-03-25
PRIOR APPLICATION NUMBER: 60/081049
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/095998
PRIOR FILING DATE: 1998-08-10
PRIOR APPLICATION NUMBER: 60/097000
PRIOR FILING DATE: 1998-08-18
PRIOR APPLICATION NUMBER: 60/099601
PRIOR FILING DATE: 1998-09-09
PRIOR APPLICATION NUMBER: 60/099803
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/099811
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/099812
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/100858
PRIOR FILING DATE: 1998-09-17
PRIOR APPLICATION NUMBER: 60/101922
PRIOR FILING DATE: 1998-09-24
PRIOR APPLICATION NUMBER: 60/106032
PRIOR FILING DATE: 1998-10-28
PRIOR APPLICATION NUMBER: 60/109304
PRIOR FILING DATE: 1998-11-20
PRIOR APPLICATION NUMBER: 60/125778
PRIOR FILING DATE: 1999-03-23
PRIOR APPLICATION NUMBER: 60/139695
PRIOR FILING DATE: 1999-06-15
PRIOR APPLICATION NUMBER: 60/145070
PRIOR FILING DATE: 1999-07-20
PRIOR APPLICATION NUMBER: 60/145698
PRIOR FILING DATE: 1999-07-26
PRIOR APPLICATION NUMBER: 60/149396
PRIOR FILING DATE: 1999-08-17
PRIOR APPLICATION NUMBER: 60/169495
PRIOR FILING DATE: 1999-12-07
PRIOR APPLICATION NUMBER: 08/918874
PRIOR FILING DATE: 1997-08-26
PRIOR APPLICATION NUMBER: 08/93821
PRIOR FILING DATE: 1997-09-19
PRIOR APPLICATION NUMBER: 08/960507
PRIOR FILING DATE: 1997-10-29
PRIOR APPLICATION NUMBER: 09/114844
PRIOR FILING DATE: 1998-07-14
PRIOR APPLICATION NUMBER: 09/136801
PRIOR FILING DATE: 1998-08-19
PRIOR APPLICATION NUMBER: 09/136804
PRIOR FILING DATE: 1998-08-19

PRIOR APPLICATION NUMBER: 09/136828
PRIOR FILING DATE: 1998-08-19
PRIOR APPLICATION NUMBER: 09/158342
PRIOR FILING DATE: 1998-09-21
PRIOR APPLICATION NUMBER: 09/180997
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 09/202088
PRIOR FILING DATE: 1998-12-08
PRIOR APPLICATION NUMBER: 09/254311
PRIOR FILING DATE: 1999-03-03
PRIOR APPLICATION NUMBER: 09/254460
PRIOR FILING DATE: 1999-03-09
PRIOR APPLICATION NUMBER: 09/254465
PRIOR FILING DATE: 1999-03-05
PRIOR APPLICATION NUMBER: 09/284663
PRIOR FILING DATE: 1999-04-15
PRIOR APPLICATION NUMBER: 09/332928
PRIOR FILING DATE: 1999-06-14
PRIOR APPLICATION NUMBER: 09/332929
PRIOR FILING DATE: 1999-06-14
PRIOR APPLICATION NUMBER: 09/333075
PRIOR FILING DATE: 1999-06-14
PRIOR APPLICATION NUMBER: 09/333077
PRIOR FILING DATE: 1999-06-14
PRIOR APPLICATION NUMBER: 09/380137
PRIOR FILING DATE: 1999-08-25
PRIOR APPLICATION NUMBER: 09/380138
PRIOR FILING DATE: 1999-08-25
PRIOR APPLICATION NUMBER: 09/380139
PRIOR FILING DATE: 1999-08-25
PRIOR APPLICATION NUMBER: 09/403296
PRIOR FILING DATE: 1999-10-18
PRIOR APPLICATION NUMBER: 09/403297
PRIOR FILING DATE: 1999-10-18
PRIOR APPLICATION NUMBER: 09/423741
PRIOR FILING DATE: 1999-11-10
PRIOR APPLICATION NUMBER: 09/423844
PRIOR FILING DATE: 1999-11-12
PRIOR APPLICATION NUMBER: 09/522342
PRIOR FILING DATE: 2000-03-09
PRIOR APPLICATION NUMBER: 09/548815
PRIOR FILING DATE: 2000-04-13
PRIOR APPLICATION NUMBER: 09/664610
PRIOR FILING DATE: 2000-09-18
PRIOR APPLICATION NUMBER: 09/665350
PRIOR FILING DATE: 2000-09-18
PRIOR APPLICATION NUMBER: 09/709238
PRIOR FILING DATE: 2001-01-08
PRIOR APPLICATION NUMBER: 09/767609
PRIOR FILING DATE: 2001-01-22
PRIOR APPLICATION NUMBER: 09/802706
PRIOR FILING DATE: 2001-03-09
PRIOR APPLICATION NUMBER: 09/808689
PRIOR FILING DATE: 2001-03-14
PRIOR APPLICATION NUMBER: 09/866028
PRIOR FILING DATE: 2001-05-25
PRIOR APPLICATION NUMBER: 09/870574
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: 09/872035
PRIOR FILING DATE: 2001-06-01
PRIOR APPLICATION NUMBER: 09/886342
PRIOR FILING DATE: 2001-06-19
PRIOR APPLICATION NUMBER: 09/US98/14552
PRIOR FILING DATE: 1998-07-14
PRIOR APPLICATION NUMBER: 09/US98/18824
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 09/US98/19093
PRIOR FILING DATE: 1998-09-14
PRIOR APPLICATION NUMBER: 09/US98/19330
PRIOR FILING DATE: 1998-09-16
PRIOR APPLICATION NUMBER: 09/US98/19437
PRIOR FILING DATE: 1998-09-17
PRIOR APPLICATION NUMBER: 09/US98/24855

```



```
; PRIOR FILING DATE: 1998-11-20
; PRIOR APPLICATION NUMBER: PCT/US98/25108
; PRIOR FILING DATE: 1998-12-01
; PRIOR APPLICATION NUMBER: PCT/US98/25190
; PRIOR FILING DATE: 1998-11-25
; PRIOR APPLICATION NUMBER: PCT/US99/05028
; PRIOR FILING DATE: 1999-03-08
; PRIOR APPLICATION NUMBER: PCT/US99/12252
; PRIOR FILING DATE: 1999-06-02
; PRIOR APPLICATION NUMBER: PCT/US99/20111
; PRIOR FILING DATE: 1999-09-01
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
```

```
Query Match 64.4%; Score 29; DB 9; Length 325;
Best Local Similarity 71.4%; Pred. No. 69;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 2 ITTNAQY 8
Db 50 LFTTAQY 56
```

```
RESULT 8
US-10-174-590-390
```

```
; Sequence 390, Application US/10174590
; Publication No. US20030008352A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Matanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C42
; CURRENT APPLICATION NUMBER: US/10/174,590
; PRIOR FILING DATE: 2002-06-18
; PRIOR APPLICATION removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 390
; LENGTH: 325
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-174-590-390
```

```
Query Match 64.4%; Score 29; DB 9; Length 325;
Best Local Similarity 71.4%; Pred. No. 69;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 2 ITTNAQY 8
Db 50 LFTTAQY 56
```

```
RESULT 9
US-10-176-758-390
```

```
; Sequence 390, Application US/10176758
; Publication No. US20030008353A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
```

```
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Matanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C104
; CURRENT APPLICATION NUMBER: US/10/176,758
; PRIOR FILING DATE: 2002-06-21
; PRIOR APPLICATION removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 390
; LENGTH: 325
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-176-758-390
```

```
Query Match 64.4%; Score 29; DB 9; Length 325;
Best Local Similarity 71.4%; Pred. No. 69;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 2 ITTNAQY 8
Db 50 LFTTAQY 56
```

```
RESULT 10
US-10-175-737-390
```

```
; Sequence 390, Application US/10175737
; Publication No. US20030013153A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Matanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C50
; CURRENT APPLICATION NUMBER: US/10/175,737
; PRIOR FILING DATE: 2002-06-19
; PRIOR APPLICATION removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 390
; LENGTH: 325
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-175-737-390
```

```
Query Match 64.4%; Score 29; DB 9; Length 325;
Best Local Similarity 71.4%; Pred. No. 69;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 2 ITTNAQY 8
Db 50 LFTTAQY 56
```

```
RESULT 11
US-09-870-574-3
```

```
; Sequence 3, Application US/09870574
; Patent No. US20020102723A1
; GENERAL INFORMATION:
; APPLICANT: Gurney, Austin L.
```

APPLICANT: Aggarwal, Sudeepa  
APPLICANT: Xie, Ming-Hong  
APPLICANT: Maruoka, Sileen M.  
APPLICANT: Foster, Jessica S.  
APPLICANT: Goddard, Audrey  
APPLICANT: Wood, William I.  
TITLE OF INVENTION: INTERLEUKIN-22 POLYPEPTIDES, NUCLEIC ACIDS ENCODING  
FILE REFERENCE: P2806-1(US)  
CURRENT APPLICATION NUMBER: US/09/870,574  
CURRENT FILING DATE: 2001-05-30  
PRIOR APPLICATION NUMBER: US 60/169,495  
PRIOR FILING DATE: 1999-12-07  
PRIOR APPLICATION NUMBER: PCT/US00/14042  
PRIOR FILING DATE: 2000-05-22  
PRIOR APPLICATION NUMBER: PCT/US00/23328  
PRIOR FILING DATE: 2000-08-24  
NUMBER OF SEQ ID NOS: 7  
SEQ ID NO 3  
LENGTH: 325  
TYPE: PRT  
ORGANISM: Homo Sapien  
US-09-870-574-3

Query Match 64.4%; Score 29; DB 10; Length 325;  
Best Local Similarity 71.4%; Pred. No. 69;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 ITFNAOY 8  
Db 50 LITFAOY 56

RESULT 12  
US-09-949-192-5  
Sequence 5, Application US/09949192  
Patent No. US20020142292A1  
GENERAL INFORMATION:  
APPLICANT: Parham, Christi L.  
APPLICANT: Gorman, Daniel L.  
APPLICANT: Kurata, Hirokazu  
APPLICANT: Arai, Naoko  
APPLICANT: Sana, Theodore R.  
APPLICANT: Mattson, Jeanine D.  
APPLICANT: Murphy, Erin E.  
APPLICANT: Savkoor, Chetan  
APPLICANT: Grein, Jeffery  
APPLICANT: Smith, Kathleen M.  
APPLICANT: McElanahan, Terrill K.  
TITLE OF INVENTION: MAMMALIAN GENES; RELATED REAGENTS AND METHODS  
FILE REFERENCE: DX01169K  
CURRENT APPLICATION NUMBER: US/09/949,192  
CURRENT FILING DATE: 2001-09-07  
PRIOR APPLICATION NUMBER: 60/231,267  
PRIOR FILING DATE: 2000-09-08  
NUMBER OF SEQ ID NOS: 53  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 5  
LENGTH: 325  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-949-192-5

Query Match 64.4%; Score 29; DB 10; Length 325;  
Best Local Similarity 71.4%; Pred. No. 69;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 ITFNAOY 8  
Db 50 LITFAOY 56

RESULT 13

US-10-052-586-390  
Sequence 390, Application US/10052586  
Patent No. US20020127584A1  
GENERAL INFORMATION:  
APPLICANT: Baker, Kevin P.  
APPLICANT: Chen, Jian  
APPLICANT: Desnoyers, Luc  
APPLICANT: Goddard, Audrey  
APPLICANT: Godowski, Paul J.  
APPLICANT: Gurney, Austin L.  
APPLICANT: Pan, James  
APPLICANT: Smith, Victoria  
APPLICANT: Watanabe, Collin K.  
APPLICANT: Wood, William I.  
APPLICANT: Zhang, Zemin  
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
FILE REFERENCE: P3430R1C1  
CURRENT APPLICATION NUMBER: US/10/052,586  
CURRENT FILING DATE: 2002-01-15  
PRIOR APPLICATION NUMBER: 60/059263  
PRIOR FILING DATE: 1997-09-18  
PRIOR APPLICATION NUMBER: 60/059266  
PRIOR FILING DATE: 1997-09-18  
PRIOR APPLICATION NUMBER: 60/062250  
PRIOR FILING DATE: 1997-10-17  
PRIOR APPLICATION NUMBER: 60/063120  
PRIOR FILING DATE: 1997-10-24  
PRIOR APPLICATION NUMBER: 60/063121  
PRIOR FILING DATE: 1997-10-24  
PRIOR APPLICATION NUMBER: 60/063486  
PRIOR FILING DATE: 1997-10-21  
PRIOR APPLICATION NUMBER: 60/063540  
PRIOR FILING DATE: 1997-10-28  
PRIOR APPLICATION NUMBER: 60/063541  
PRIOR FILING DATE: 1997-10-28  
PRIOR APPLICATION NUMBER: 60/063544  
PRIOR FILING DATE: 1997-10-28  
PRIOR APPLICATION NUMBER: 60/063564  
PRIOR FILING DATE: 1997-10-28  
PRIOR APPLICATION NUMBER: 60/063734  
PRIOR FILING DATE: 1997-10-29  
PRIOR APPLICATION NUMBER: 60/063870  
PRIOR FILING DATE: 1997-10-31  
PRIOR APPLICATION NUMBER: 60/064103  
PRIOR FILING DATE: 1997-10-31  
PRIOR APPLICATION NUMBER: 60/065311  
PRIOR FILING DATE: 1997-11-13  
PRIOR APPLICATION NUMBER: 60/065120  
PRIOR FILING DATE: 1997-11-21  
PRIOR APPLICATION NUMBER: 60/066466  
PRIOR FILING DATE: 1997-11-24  
PRIOR APPLICATION NUMBER: 60/066772  
PRIOR FILING DATE: 1997-11-24  
PRIOR APPLICATION NUMBER: 60/069335  
PRIOR FILING DATE: 1997-12-11  
PRIOR APPLICATION NUMBER: 60/069425  
PRIOR FILING DATE: 1997-12-12  
PRIOR APPLICATION NUMBER: 60/069870  
PRIOR FILING DATE: 1997-12-17  
PRIOR APPLICATION NUMBER: 60/068017  
PRIOR FILING DATE: 1997-12-18  
PRIOR APPLICATION NUMBER: 60/077450  
PRIOR FILING DATE: 1998-03-10  
PRIOR APPLICATION NUMBER: 60/077632  
PRIOR FILING DATE: 1998-03-11  
PRIOR APPLICATION NUMBER: 60/077649  
PRIOR FILING DATE: 1998-03-11  
PRIOR APPLICATION NUMBER: 60/078886  
PRIOR FILING DATE: 1998-03-20  
PRIOR APPLICATION NUMBER: 60/078939  
PRIOR FILING DATE: 1998-03-20  
PRIOR APPLICATION NUMBER: 60/079664

PRIOR FILING DATE: 1998-03-27  
 PRIOR APPLICATION NUMBER: 60/079786  
 PRIOR FILING DATE: 1998-03-27  
 PRIOR APPLICATION NUMBER: 60/080107  
 PRIOR FILING DATE: 1998-03-31  
 PRIOR APPLICATION NUMBER: 60/080194  
 PRIOR FILING DATE: 1998-03-31  
 PRIOR APPLICATION NUMBER: 60/080327  
 PRIOR FILING DATE: 1998-04-01  
 PRIOR APPLICATION NUMBER: 60/080333  
 PRIOR FILING DATE: 1998-04-01  
 PRIOR APPLICATION NUMBER: 60/081049  
 PRIOR FILING DATE: 1998-04-08  
 PRIOR APPLICATION NUMBER: 60/081070  
 PRIOR FILING DATE: 1998-04-08  
 PRIOR APPLICATION NUMBER: 60/081195  
 PRIOR FILING DATE: 1998-04-09  
 PRIOR APPLICATION NUMBER: 60/081838  
 PRIOR FILING DATE: 1998-04-15  
 PRIOR APPLICATION NUMBER: 60/082568  
 PRIOR FILING DATE: 1998-04-21  
 PRIOR APPLICATION NUMBER: 60/082569  
 PRIOR FILING DATE: 1998-04-21  
 PRIOR APPLICATION NUMBER: 60/082704  
 PRIOR FILING DATE: 1998-04-22  
 PRIOR APPLICATION NUMBER: 60/082797  
 PRIOR FILING DATE: 1998-04-22  
 PRIOR APPLICATION NUMBER: 60/083322  
 PRIOR FILING DATE: 1998-04-28  
 PRIOR APPLICATION NUMBER: 60/083495  
 PRIOR FILING DATE: 1998-04-29  
 PRIOR APPLICATION NUMBER: 60/083496  
 PRIOR FILING DATE: 1998-04-29  
 PRIOR APPLICATION NUMBER: 60/083499  
 PRIOR FILING DATE: 1998-04-29  
 PRIOR APPLICATION NUMBER: 60/083559  
 PRIOR FILING DATE: 1998-04-29  
 PRIOR APPLICATION NUMBER: 60/084366  
 PRIOR FILING DATE: 1998-05-05  
 PRIOR APPLICATION NUMBER: 60/084414  
 PRIOR FILING DATE: 1998-05-06  
 PRIOR APPLICATION NUMBER: 60/084639  
 PRIOR FILING DATE: 1998-05-07  
 PRIOR APPLICATION NUMBER: 60/084640  
 PRIOR FILING DATE: 1998-05-07  
 PRIOR APPLICATION NUMBER: 60/084643  
 PRIOR FILING DATE: 1998-05-07  
 PRIOR APPLICATION NUMBER: 60/085573  
 PRIOR FILING DATE: 1998-05-15  
 PRIOR APPLICATION NUMBER: 60/085579  
 PRIOR FILING DATE: 1998-05-15  
 PRIOR APPLICATION NUMBER: 60/085580  
 PRIOR FILING DATE: 1998-05-15  
 PRIOR APPLICATION NUMBER: 60/085582  
 PRIOR FILING DATE: 1998-05-15  
 PRIOR APPLICATION NUMBER: 60/085700  
 PRIOR FILING DATE: 1998-05-15  
 PRIOR APPLICATION NUMBER: 60/086023  
 PRIOR FILING DATE: 1998-05-18  
 PRIOR APPLICATION NUMBER: 60/086392  
 PRIOR FILING DATE: 1998-05-22  
 PRIOR APPLICATION NUMBER: 60/086486  
 PRIOR FILING DATE: 1998-05-22  
 PRIOR APPLICATION NUMBER: 60/087098  
 PRIOR FILING DATE: 1998-05-28  
 PRIOR APPLICATION NUMBER: 60/087208  
 PRIOR FILING DATE: 1998-05-28  
 PRIOR APPLICATION NUMBER: 60/087609  
 PRIOR FILING DATE: 1998-06-02  
 PRIOR APPLICATION NUMBER: 60/087759  
 PRIOR FILING DATE: 1998-06-02  
 PRIOR APPLICATION NUMBER: 60/087827  
 PRIOR FILING DATE: 1998-06-03

PRIOR APPLICATION NUMBER: 60/088025  
 PRIOR FILING DATE: 1998-06-04  
 PRIOR APPLICATION NUMBER: 60/088028  
 PRIOR FILING DATE: 1998-06-04  
 PRIOR APPLICATION NUMBER: 60/088029  
 PRIOR FILING DATE: 1998-06-04  
 PRIOR APPLICATION NUMBER: 60/088033  
 PRIOR FILING DATE: 1998-06-04  
 PRIOR APPLICATION NUMBER: 60/088167  
 PRIOR FILING DATE: 1998-06-05  
 PRIOR APPLICATION NUMBER: 60/088202  
 PRIOR FILING DATE: 1998-06-05  
 PRIOR APPLICATION NUMBER: 60/088212  
 PRIOR FILING DATE: 1998-06-05  
 PRIOR APPLICATION NUMBER: 60/088217  
 PRIOR FILING DATE: 1998-06-05  
 PRIOR APPLICATION NUMBER: 60/088326  
 PRIOR FILING DATE: 1998-06-04  
 PRIOR APPLICATION NUMBER: 60/088655  
 PRIOR FILING DATE: 1998-06-09  
 PRIOR APPLICATION NUMBER: 60/088722  
 PRIOR FILING DATE: 1998-06-10  
 PRIOR APPLICATION NUMBER: 60/088738  
 PRIOR FILING DATE: 1998-06-10  
 PRIOR APPLICATION NUMBER: 60/088740  
 PRIOR FILING DATE: 1998-06-10  
 PRIOR APPLICATION NUMBER: 60/088811  
 PRIOR FILING DATE: 1998-06-10  
 PRIOR APPLICATION NUMBER: 60/088824  
 PRIOR FILING DATE: 1998-06-10  
 PRIOR APPLICATION NUMBER: 60/088825  
 PRIOR FILING DATE: 1998-06-10  
 PRIOR APPLICATION NUMBER: 60/088826  
 PRIOR FILING DATE: 1998-06-10  
 PRIOR APPLICATION NUMBER: 60/088861  
 PRIOR FILING DATE: 1998-06-11  
 PRIOR APPLICATION NUMBER: 60/088863  
 PRIOR FILING DATE: 1998-06-11  
 PRIOR APPLICATION NUMBER: 60/088876  
 PRIOR FILING DATE: 1998-06-11  
 PRIOR APPLICATION NUMBER: 60/089090  
 PRIOR FILING DATE: 1998-06-12  
 PRIOR APPLICATION NUMBER: 60/089105  
 PRIOR FILING DATE: 1998-06-12  
 PRIOR APPLICATION NUMBER: 60/089512  
 PRIOR FILING DATE: 1998-06-16  
 PRIOR APPLICATION NUMBER: 60/089514  
 PRIOR FILING DATE: 1998-06-16  
 PRIOR APPLICATION NUMBER: 60/089538  
 PRIOR FILING DATE: 1998-06-17  
 PRIOR APPLICATION NUMBER: 60/089598  
 PRIOR FILING DATE: 1998-06-17  
 PRIOR APPLICATION NUMBER: 60/089653  
 PRIOR FILING DATE: 1998-06-17  
 PRIOR APPLICATION NUMBER: 60/089908

Query Match 64.4%; Score 29; DB 12; Length 325;  
 Best Local Similarity 71.4%; Pred. No. 69;  
 Matches 5; Conservative 1; Mismatches 0; Gaps 0;  
 QY 2 ITPNAY 8  
 Db 50 LITTAAY 56  
 RESULT 14  
 US-09-925-297-684  
 Sequence 684, Application US/09925297  
 Patent No. US20020081659A1  
 GENERAL INFORMATION:  
 APPLICANT: Rosen et al.  
 TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
 FILE REFERENCE: PA105

; CURRENT APPLICATION NUMBER: US/09/925,297  
; CURRENT FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: PCT/US00/05989  
; PRIOR FILING DATE: 2000-03-08  
; PRIOR APPLICATION NUMBER: 60/124,270  
; PRIOR FILING DATE: 1999-03-12  
; NUMBER OF SEQ ID NOS: 928  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 684  
; LENGTH: 427  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; US-09-925-297-684

Query Match 64.4%; Score 29; DB 10; Length 427;  
Best Local Similarity 100.0%; Pred. No. 93;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AITFNA 6  
|||  
Db 172 AITFNA 177

RESULT 15  
US-09-801-368-244  
; Sequence 244, Application US/09801368  
; Patent No. US20020128250A1  
; GENERAL INFORMATION:  
; APPLICANT: Busby, Robert  
; APPLICANT: Cali, Brian  
; APPLICANT: Hecht, Peter  
; APPLICANT: Holtzman, Doug  
; APPLICANT: Madden, Kevin  
; APPLICANT: Maxon, Mary  
; APPLICANT: Milne, Todd  
; APPLICANT: No. US20020128250A1man, Thea  
; APPLICANT: Royer, John  
; APPLICANT: Salama, Sofie  
; APPLICANT: Sherman, Amir  
; APPLICANT: Silva, Jeff  
; APPLICANT: Summers, Eric  
; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi  
; FILE REFERENCE: 109272.147  
; CURRENT APPLICATION NUMBER: US/09/801,368  
; CURRENT FILING DATE: 2001-03-07  
; PRIOR APPLICATION NUMBER: US 09/487,558  
; PRIOR FILING DATE: 2000-01-19  
; PRIOR APPLICATION NUMBER: US 60/160,587  
; PRIOR FILING DATE: 1999-10-20  
; NUMBER OF SEQ ID NOS: 440  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 244  
; LENGTH: 1564  
; TYPE: PRT  
; ORGANISM: Saccharomyces cerevisiae  
; US-09-801-368-244

Query Match 64.4%; Score 29; DB 10; Length 1564;  
Best Local Similarity 62.5%; Pred. No. 3.9e+02;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 AITFNAOY 8  
|||  
Db 635 AITFNAOY 642

Search completed: January 29, 2003, 14:04:56  
Job time : 7.92308 secs